1c

PRACTICAL PROBLEMS IN STATISTICS

By

D. N. ELHANCE, M. COM.,

an, Faculty of Commerce, University of Jodhpur
Author of "Principles of Statistics" (in Hindi)
"Fundamentals of Statistics", etc.

Revised by

DR. PRASAN PRAKASH, M. A., Ph. D.,

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

Seventeenth Revised and Enlarged Edition

KITAB MAHAL, ALLAHABAD.
1977

First Edition, 1952
Seventeenth Edition, 1977

(Binding)

S.C.E.R.T., West Bengal
Date 15.3.78
Acc. No. 2652

311
ELH

Published by :—Kitab Mahal, 15, Thornhill Road, Allahabad.
Printed by :—Kitab Mahal, (W. D.) Pvt. Ltd., Allahabad.

PREFACE TO THE SEVENTEENTH EDITION

The book completed sixteen editions within the period of annual le shows its popularity and overall usefulness and we are glad note that the previous editions served the objectives for which is book was written.

The present edition is a thoroughly revised and enlarged edition ve have added 20 problems and 217 revisional problems in the t thereby increasing the problems from 729 to 966. The introtion of revisional problems along with the answers to these blems will serve a very useful purpose. It will enable the rs, students and practitioners of statistics to solve problems out any assistance and to verify their answers with the ans- s calculated by us. Another feature of the present edition hat all the questions have been up-dated and have been made e realistic in terms of price, height, weight, and other variables all the solutions have been checked and rechecked to ascer- r. their correctness. We have sincerely endeavoured to remove misprints and mistakes from this edition to make it more de- dable for the solutions to the problems.

We have tried to incorporate valuable suggestions received n colleagues and students as far as it had been possible for us do so.

The present edition has been printed by using a different method rinting than the previous editions to avoid mistakes that occur he printing stage. The presentation of the present edition is sing to the eye and is very attractive. The system of num- g problems has also been changed and has been modernised. each problem begins with the number of that chapter and sing decimals the next number indicates the problem number problem no. 5 in chapter third would be numbered as 3.5.

It is sincerely hoped that the present edition would be found to ore useful by the students and other users than the previous ns. The present thoroughly revised and considerably en- d edition will, it is hoped, serve its purpose in fields wider before.

D. N. ELHANCE

PREFACE TO THE FIRST EDITION

It is impossible to overemphasise the importance of statis
in modern times. In fact modern age is the age of statistics,
all branches of human knowledge make extensive use of st
tical methods. In our country also the importance of statis
has been realised though quite late, and practically all the uni
sities are now imparting training in statistical methods and
thus trying to fill up the ever-increasing demand of statistici

However, as a teacher of statistics for a long time, I have
that students need one such book on the subjec[illegible] which
can solve practical questions and find out the [illegible]cy or ot
wise of their answers. No doubt, there are many book
foreign as well as Indian authors which give some question
the end of each chapter for practice, but there is no one boo
which all possible types of questions to important topics
be found at a place. This book is an attempt to meet this
culty. Questions have been taken from the examination pa
of all the universities in the country and also from P. C. S.
I. A. S. examination papers. Many new types of questions
been framed by me and the book contains problems of practic
all possible varieties. I am sure the book would be of imme
help to students of statistics in the universities as also to t
who offer [illegible] subject in competitive examinations.

The list of topics covered is an exhaustive one and hal
book is devoted to advanced problems on probability, sam
chi-square test, analysis of variance, etc.

I have been assisted in my work by my old student
T. N. Bhargava, M. Com., and I am grateful to him for th
would be highly thankful to those who would kindly poi
the inaccuracies in the book and give suggestions for im
ment.

1st June, 1952 D. N. ELH

Allahabad.

CONTENTS

	Problems	Pages
CHAPTER 1		
Measures of Central Tendency		
Mean	1.1 — 1.19	1— 16
Weighted Mean	1.20— 1.27	16— 21
Median	1.28— 1.45	21— 37
Mode	1.46— 1.65	38— 52
Geometric Mean	1.66— 1.75	52— 58
Harmonic Mean	1.76— 1.87	58— 64
Revisional Problems-1	1.88— 1.107	64— 67
CHAPTER 2		
Measures of Dispersion		
Range	2.1 — 2.3	68— 70
Quartile Deviation	2.4 — 2.7	70— 72
Average Deviation	2.8 — 2.17	73— 83
Standard Deviation	2.18— 2.51	84—119
Revisional Problems-2	2.52— 2.69	119—122
CHAPTER 3		
Moments, Kurtosis and Skewness		
Moments, Kurtosis and Skewness ...	3.1 — 3.25	123—157
Revisional Problems-3	3.26— 3.34	157—158
CHAPTER 4		
Correlation		
Correlation	4.1 — 4.37	159—209
Revisional Problems-4	4.38— 4.47	209—210
CHAPTER 5		
Regression		
Regression	5.1 — 5.13	211—230
Revisional Problems-5	5.14— 5.25	230—232
CHAPTER 6		
Index Numbers		
Index Numbers	6.1 — 6.24	233—252
Revisional Problems-6	6.25— 6.34	252—253

CHAPTER 7

Analysis of Time Series

		Problems	Pages
Analysis of Time Series	...	7.1 — 7.19	254—273
Revisional Problems-7 ...	...	7.20— 7.29	273—274

CHAPTER 8

Interpolation

		Problems	Pages
Parabolic Curve Method	...	8.1 — 8.5	275—280
Newton's Method ...	...	8.6 — 8.23	280—299
Newton Gauss and Sterling Formula ...	...	8.24— 8.25	299—301
Sterling Formula ...	...	8.26— 8.27	301—303
Newton Gauss Backward Formula ...	...	8.28— 8.29	303—304
Binominal Expansion Method	...	8.30— 8.42	304—313
Lagrange's Method ...	...	8.43— 8.51	314—321
Divided Difference Method	...	8.52— 8.56	322—325
Revisional Problems-8 ...	...	8.57— 8.65	325—326

CHAPTER 9

Association of Attributes

		Problems	Pages
Classification of Data ...	...	9.1 — 9.12	327—333
Association ...	...	9.13— 9.31	334—345
Revisional Problems-9 ...	...	9.32— 9.41	346—347

CHAPTER 10

Consistence of Data

		Problems	Pages
Consistence of Data ...	...	10.1 —10.21	348—357
Revisional Problems-10 ...	..	10.22—10.31	357—358

CHAPTER 11

Probability

		Problems	Pages
Probability ...	...	11.1 —11.99	359—39
Revisional Problems-11	...	11.100—11.115	39

CHAPTER 12

Probability (Contd.)

		Problems	Pages
Mathematical Treatment of Probability ...	...	12.1 —12.23	392—40
Mathematical Expectation	...	12.24—12.28	408—41
Revisional Problems-12	...	12.29—12.40	412—41

CHAPTER 13

Theoretical Frequency Distributions

	Problems	Pages
Binomial Distribution	13.1 —13.10	414—422
Poisson Distribution	13.11—13.24	423—433
Normal Distribution	13.25—13.39	433—443
Revisional Problems-13	13.40—13.50	443—444

CHAPTER 14

Sampling of Attributes—Large Samples

	Problems	Pages
Standard Deviation (of numbers)	14.1 —14.8	445—448
Standard Deviation (of proportions)	14.9 —14.21	448—454
Standard Error (of the (differences)	14.22—14.34	454—461
Revisional Problems-14	14.35—14.44	461—462

CHAPTER 15

Sampling of Variables—Large Samples

	Problems	Pages
Standard Error of the Mean	15.1 —15.15	463—471
Standard Error of Median, Mean Deviation, Standard Deviation, etc.	15.16—15.26	471—477
Standard Error of C. v. 'R' Reg. etc.	15.26—15.36	477—483
Revisional Problems-15	15.37—15.46	483—484

CHAPTER 16

Sampling of Variables—Large Samples (Contd.)

	Problems	Pages
Sampling of Variables—Large Samples (Contd.)	16.1 —16.27	485—495
Revisional Problems-16	16.28—16.37	495—496

HAPTER 17

X^2 (Chi Square), Test and Co-efficient of Contingency

	Problems	Pages
X^2 (Chi-Square) Test	17.1 —17.38	497—525
X^2 and Normal Distribution	17.39—17.40	526—527
Co-efficient of Contingency	17.41—17.46	527—533
Revisional Problems-17	17.47—17.56	534—535

CHAPTER 18

Sampling of Variables—Small Samples

Significance of Mean ... 18.1 —18.19 536—5
Significance of the Difference of Means ... ... 18.20—18.40 554—5
Significance of 'r' ... ... 18.41—18.57 575—5
Revisional Problems-18 ... 18.58—18.67 583—5

CHAPTER 19

Analysis of Variance

Analysis of Variance ... ... 19.1 —19.19 585—6
Revisional Problems-19 . . 19.20—19.29 606—6
Mathematical Tables 609—6

asures of Central Tendency

I

Measures of Central Tendency or averages are representatives which are to indicate the value of large group of numbers. There are five kinds of ges in common use. They are (1) Arithmetic Average or Mean, (2) Median, Aode, (4) Geometric Mean and (5) Harmonic Mean. In the present chapter ave dealt with various types of problems connected with these five measures ntral tendency.

First of all the problems connected with the calculation of Mean have solved. Mean or Arithmetic Average is obtained by adding the values of s and dividing the total by the number of items.

After this we have dealt with the problems connected with Median. ian is the value of the middle item in a series. To find out the median a series st arranged in ascending or descending order and then the value of the middle is found out either by mere observation or by interpolation.

Next comes Mode. Mode is the most popular item of the series. The e which occurs the largest number of times in a series is known as Mode. again can be found out either by observation or by interpolation.

Geometric Mean has been dealt with after Mode. Geometric Mean is the root of the product of *n* items. Its calculation involves the use of logarithm es.

Finally we have dealt with problems of Harmonic Mean. Harmonic an is the reciprocal of the arithmetic mean of the reciprocals of the numbers aged.

In all these averages first we have solved some questions where the series sist of individual observations only, and then questions of discrete series and tinuous series have been solved.

MEAN

PROBLEM 1.1. *The following table gives the monthly income of twelve families town :*

l: No.	*Monthly Income*	*Sl. No.*	*Monthly Income*
	Rs.		Rs.
1	*280*	*7*	*80*
2	*180*	*8*	*84*
3	*96*	*9*	*100*
4	*98*	*10*	*76*
5	*104*	*11*	*600*
6	*75*	*12*	*20*

Calculate the arithmetic average of the above income.

SOLUTION. Computation of arithmetic average.

Income (m)
280
180
96
98
104
75
80
84
100
76
600
20

$\Sigma m = 1793$

Arithmetic average or $a = \frac{\Sigma m}{n}$, where Σm represents summation of measurements and n, the number of items.

$$a = \frac{1793}{12} = \text{Rs. } 149.42 \text{ approx.}$$

PROBLEM 1.2. *Calculate the simple arithmetic average of the followin items :—*

Size of items

20
28
34
39
42
50
53
54
59
64
72
74
75
78
79

SOLUTION. Computation of arithmetic average..

Size of items (m)
20
28
34
39
42
50
53
54
59
64
72
74
75
78
79

$\Sigma m = 821$

Arithmetic average or $a = \frac{\Sigma m}{n}$, where Σm represents summation of measurements and n, the number of items.

$$a = \frac{821}{15} = 54{\cdot}73$$

Arithmetic average of the series= 54·73.

Alternative Method :

(*m*)	Deviation from an assumed mean (50) (*dx*)
20	—30
28	—22
34	—16
39	—11
42	— 8
50	0
53	3
54	4
59	9
64	14
72	22
74	24
75	25
78	28
79	29
$n=15$	$\Sigma dx=+71$

Arithmetic average or $a=x+\frac{\Sigma dx}{n}$; where x represents assumed average and Σdx represents the total of the deviations from the assumed average and n, the number of items.

$$a=50+\frac{71}{15}=50+4{\cdot}73=54{\cdot}73$$

Arithmetic average of the series=54.73

PROBLEM 1.3. *Find out the mean height from the following data relating to height measurements of 8 persons in centimetres.*

159, 161, 163, 165, 167, 169, 171, 173

SOLUTION. Computation of mean height

Height (*m*)	*Deviations from* as. av. (167) (*dx*)	*Step deviations* (*i*) (*dx′*)
159	—8	—4
161	—6	—3
163	—4	—2
165	—2	—1
167	0	0
169	2	1
171	4	2
173	6	3
$n=8$		$\Sigma dx'=-4$

Arithmetic average or $a=x+\left(\frac{\Sigma dx'\times i}{n}\right)$; where $\Sigma dx'$ represents the ummation of step deviations and '*i*' represents the common factor (class inter- al) of deviation.

$$a=167+\left(\frac{-4}{8}\times 2\right)=167-1=166 \text{ Cm.}$$

Mean height=166 Cm.

PROBLEM 1.4. *The following table gives the number of children bor per family in 735 families. Calculate the average number of children born pe family.*

Number of children born per family	*Number of families*
0	*96*
1	*108*
2	*154*
3	*126*
4	*95*
5	*62*
6	*45*
7	*20*
8	*11*
9	*6*
10	*5*
11	*5*
12	*1*
13	*1*

SOLUTION. Computation of the average number of children born pe family.

Number of children born per family (*m*)	*Number of families* (*f*)	*m× f* (*mf*)
0	96	0
1	108	108
2	154	308
3	126	378
4	95	380
5	62	310
6	45	270
7	20	140
8	11	88
9	6	54
10	5	50
11	5	55
12	1	12
13	1	13
	$\Sigma f = 735$	$\Sigma mf = 2166$

Arithmetic average or $a=\frac{\Sigma mf}{\Sigma f}$, where *mf* represents the product of size of item and frequency and *f* represents frequency.

$$a=\frac{2166}{735}=2.9=3 \text{ children approx.}$$

The average number of children born per family is equal to 3 appro mately.

PROBLEM 1.5. *The following data give an actual distribution, obta by tossing ten coins 1024 times and recording the number of heads that appe on each toss. What is the average number of heads per toss ?*

Number of heads	*Frequency*
0	*1*
1	*16*
2	*42*
3	*126*
4	*199*
5	*253*
6	*209*
7	*118*
8	*53*
9	*4*
10	*3*

SOLUTION. Computation of average number of heads per toss.

Number of heads (m)	Frequency (f)	($m \times f$)
0	1	0
1	16	16
2	42	84
3	126	378
4	199	796
5	253	1265
6	209	1254
7	118	826
8	53	424
9	4	36
10	3	30
	$n=1024$	$\Sigma mf=5109$

Arithmetic average $a=\frac{\Sigma mf}{n}=\frac{5109}{1024}=5$ heads per toss approx.

Thus the average number of heads per toss is 5.

PROBLEM 1·6. *What are the desiderata for a satisfactory average ? Point out the special characteristics of the arithmetic mean, the median and the geometrical mean.*

Explain the step-deviation method for finding out the arithmetic mean of a frequency distribution. Derive the usual formula and apply it to find the arithmetic mean of the distribution :

Variate—5, 10, 15, 20, 25, 30, 35, 40, 45, 50.

Frequency—20, 43, 75, 67, 72, 45, 39, 9, 8, 6.

SOLUTION. Calculation of arithmetic average by step-deviation method.

Variate (m)	Frequency (f)	Step Deviations for as. mean. 25 (dx)	Total Deviations (fdx)
5	20	-4	-80
10	43	-3	-129
15	75	-2	-150
20	67	-1	-67
25	72	0	0
30	45	$+1$	$+45$
35	39	$+2$	$+78$
40	9	$+3$	$+27$
45	8	$+4$	$+32$
50	6	$+5$	$+30$
	$n=384$		$\Sigma fdx=-214$

$$\text{Arithmetic average} = x\left(\frac{\Sigma fdx}{n} \times i\right) = 25 + \left(\frac{-214}{384} \times 5\right)$$

$$= 25 - 2.8 = 22.2$$

PROBLEM. 1.7. *Given the following frequency distribution, calculate t Arithmetic Average.*

Monthly Wages *Rs. Rs.*	*Workers*	*Monthly Wages* *Rs. Rs.*	*Workers*
212.50—217.50	*2*	*237.50—242.50*	*4*
217.50—222.50	*22*	*242.50—247.50*	*6*
222.50—227.50	*19*	*247.50—252.50*	*1*
227.50—232.50	*14*	*252.50—257.50*	*1*
232.50—237.50	*3*		

SOLUTION. Calculation of the arithmetic average.

Monthly Wages (*m*)	Mid-values of the groups (*mv*)	Workers (*f*)	Deviations from the as. av. (230) *dx*	Total Devi tions *fdx*
Rs. Rs.				
212.50—217.50	215	2	−15	−30
217.50—222.50	220	22	−10	−220
222.50—227.50	225	19	− 5	− 95
227.50—232.50	230	14	0	
232.50—237.50	235	3	+ 5	+ 15
237.50—242.50	240	4	+10	+ 40
242.50—247.50	245	6	+15	+ 90
247.50—252.50	250	1	+20	+ 20
252.50—257.50	255	1	+25	+ 25
		n=72		Σ*fdx*=−155

Applying the short-cut method.

$$\text{Arithmetic Average} = x + \frac{\Sigma dx}{n} = 230 + \frac{-155}{72}$$

$$= 230 - 2.15 = 227.85 \text{ rupees}$$

Thus the average wage of the workers is Rs. 227.85.

PROBLEM 1.8. *The frequency distribution of cost of production of Gur i rupees per quintal for different holdings in two districts is given below. Find th average cost in each district, and test whether there is any significant difference.*

Cost in rupees per quintal	*District A*	*District B*
100-102	*9*	*1*
102-104	*32*	*10*
104-106	*37*	*34*
106-108	*21*	*23*
108-110	*13*	*21*
110-112	*7*	*14*
112-114	*5*	*10*
114-116	*2*	*9*
116-118	*1*	*5*
118-120	*2*	*2*
120-122	*1*	*1*
Total	*130*	*130*

SOLUTION. Computation of the average cost of production of Gur in Districts A and B.

Cost in rupees per quintal (m)	Mid-values ($m.\ v.$)	Deviation from the as. av. (107) dx	District A		District B	
			Frequency (f)	Total deviations fdx	Frequency (f)	Total deviations fdx
100-102	101	-6	9	-54	1	-6
102-104	103	-4	32	-128	10	-40
104-106	105	-2	37	-74	34	-68
106-108	107	0	21	0	23	0
108-110	109	$+2$	13	$+26$	21	$+42$
110-112	111	$+4$	7	$+28$	14	$+56$
112-114	113	$+6$	5	$+30$	10	$+60$
114-116	115	$+8$	2	$+16$	9	$+72$
116-118	117	$+10$	1	$+10$	5	$+50$
118-120	119	$+12$	2	$+24$	2	$+24$
120-122	121	$+14$	1	$+14$	1	$+14$
			$n=130$	$\Sigma fdx=-108$	$n=130$	$\Sigma fdx=+204$

In District A, the average cost per quintal is Rs. 106.17

$$a=x+\frac{\Sigma fdx}{n}=107+\frac{-108}{130}$$

$$=107-.83=106.17$$

In District B, the average cost per quintal is Rs. 108.57

$$a=107+\frac{204}{130}$$

$$=107+1.57=108.57$$

PROBLEM 1.9 *The following table gives the frequency distribution of the percentage weekly gains or losses of a random forecast record, divided by one-half the corresponding stock market gain or loss, for the 230 weeks from January 1, 1971 to June 1, 1973.*

Percentage Gain or Loss	*Frequency*
96.50- 97.49%	*1*
97.50- 98.49%	*13*
98.50- 99.49%	*32*
99.50-100.49%	*115*
100.50-101.49%	*45*
101.50-102.49%	*14*
102.50-103.49%	*2*
103.50-104.49%	*4*
104.50-105.49%	*2*
105.50-106.49%	*1*
106.50-107.49%	*1*
Total	*230*

What is the average gain or loss for this distribution ?

SOLUTION. Computation of average gain or loss.

Percentage gain or loss	Mid points (m)	Frequency (f)	Deviations from the assumed average (100%) (dx)	Total deviations (fdx)
96.50- 97.49%	97%	1	−3	− 3
97.50- 98.49%	98%	13	−2	−26
98.50- 99.49%	99%	32	−1	−32
99.50-100.49%	100%	115	0	0
100.50-101.49%	101%	45	+1	+45
101.50-102.49%	102%	14	+2	+28
102.50-103.49%	103%	2	+3	+ 6
103.50-104.49%	104%	4	+4	+16
104.50-105.49%	105%	2	+5	+10
105.50-106.49%	106%	1	+6	+ 6
106.50-107.49%	107%	1	+7	+ 7
		$n = 230$		$\Sigma fdx = +57$

Applying the short-cut method :

$$a = 100 + \frac{57}{230} = 100 + {\cdot}25 = 100.25\% \text{ approx.}$$

Therefore, the average percentage of gain or loss for the given distribution is 100.25.

PROBLEM. 1.10. *The following table gives the population of males at different age-groups of the U. K. and India at the time of a census.*

Age-group	*U. K.(Lakhs)*	*India(Lakhs)*
0- 5	*18*	*214*
5-10	*19*	*258*
10-15	*20*	*222*
15-20	*18*	*157*
20-25	*16*	*145*
25-30	*14*	*161*
30-40	*27*	*257*
40-50	*25*	*184*
50-60	*19*	*120*
Above 60	*17*	*100*

Compare the average age of males in the two countries, and account for the difference, if any.

SOLUTION. Computation of the average age of males in the U. K. and India.

Age group (m)	Mid-values ($m.v.$)	Deviations from the as. av. (dx) (27.5)	U. K.		India	
			Population of males in lakhs (f)	Total Deviations (fdx)	Population of males in lakhs (f)	Total Deviations (fdx)
0- 5	2.5	−25	18	−450	214	−5350
5-10	7.5	−20	19	−380	258	−5160
10-15	12.5	−15	20	−300	222	−3330
15-20	17.5	−10	18	−180	157	−1570
20-25	22.5	− 5	16	− 80	145	− 725
25-30	27.5	0	14	0	161	0
30-40	35.0	+ 7.5	27	+202.50	+257	+1927.50
40-50	45.0	+17.5	25	+437.50	+184	+3220.00
50-60	55.0	+27.5	19	+522.50	+120	+3300.00
Above 60	65.0	+37.5	17	+637.50	+100	+3750.
			$n=193$	$\Sigma fdx=$ +410.00	$n=1818$	$\Sigma fdx=$ −3937.50

The average age of males in the U.K. is 29.6 years as

$$a = x + \frac{\Sigma fdx}{n} = 27.5 + \frac{410}{193} = 27.5 + 2.12 = 29.62$$

The average age of males in India is 25.33 years as

$$a = x + \frac{\Sigma fdx}{n} = 27.5 + \frac{-3937.5}{1818} = 27.5 - 2.17 = 25.33$$

Thus, the average age of males in U.K. is higher than the average age of males in India.

PROBLEM 1.11. *The following table gives the male population of Kanpur and Jaipur in 1971 :—*

Age group (years)	*Population of males in thousands*	
	Kanpur	*Jaipur*
0- 5	*14*	*9*
5-10	*13*	*8*
10-15	*13*	*8*
15-20	*13*	*7*
20-30	*33*	*15*
30-40	*29*	*12*
40-50	*17*	*9*
50-60	*7*	*6*
60-80	*4*	*4*

Calculate the average age of males at Kanpur and Jaipur separately, and account for the difference, if any.

SOLUTION. Computation of average age of males at Kanpur and Jaipur.

Age group (m)	Mid-values (m.v.)	Deviations from as. av. (25)	Step Deviations (dx)	KANPUR		JAIPUR	
				Population (000) (f)	(fdx)	Population (000) (f)	(fdx)
00-05	2.5	−22.5	−4.5	14	−63.0	9	−40.5
05-10	7.5	−17.5	−3.5	13	−45.5	8	−28.0
10-15	12.5	−12.5	−2.5	13	−32.5	8	−20.0
15-20	17.5	− 7.5	−1.5	13	−19.3	7	−10.5
20-30	25.0	0	0	33	0	15	
30-40	35.0	+10	+2	29	+58.0	12	+24.0
40-50	45.0	+20	+4	17	+68.0	9	+36.0
50-60	55.0	+30	+6	7	+42.0	6	+36.0
60-80	70.0	+45	+9	4	+36.0	4	+36.0
				$n=143$	$\Sigma fdx=$ +43.5	$n=78$	$\Sigma fdx=$ +33.0

(1) Average age of males at Kanpur :—

$$a = x + \left(\frac{\Sigma fdx}{n} \times i\right)$$

$$= 25 + \left(\frac{43.5}{143} \times 5\right)$$

$$= 25 + 1.5 = 26.5 \text{ years}$$

(2) Average of males at Jaipur :—

$$a=x+\left(\frac{\Sigma fdx}{n}\times i\right)$$

$$=25+\left(\frac{33}{78}\times 5\right)$$

$$=25+2.1=27.1 \text{ years}$$

Thus the average age of the population of Jaipur is higher than the averag age of the population of Kanpur.

PROBLEM 1.12. *Find the average marks of a student from the followin table.*

Below	*Marks*	*No. of Students*
,,	*10*	*25*
,,	*20*	*40*
,,	*30*	*60*
,,	*40*	*75*
,,	*50*	*95*
,,	*60*	*125*
,,	*70*	*190*
,,	*80*	*240*

SOLUTION. This cumulative series should be turned into a simple one in order to find out the average.

Calculation of average marks of 240 students.

Marks (*m*)	No. of Students (*f*)	Mid-Values (*m.v.*)	Deviation from assumed average (45)	Step Deviation (*dx*)	Total Deviations (*fdx*)
0-10	25	5	−40	−4	−100
10-20	15	15	−30	−3	− 45
20-30	20	25	−20	−2	− 40
30-40	15	35	−10	−1	− 15
40-50	20	45	0	0	0
50-60	30	55	+10	+1	+ 30
60-70	65	65	+20	+2	+130
70-80	50	75	+30	+3	+150
	$n=240$				$\Sigma fdx=$ +110

$$a-x+\left(\frac{\Sigma fdx}{n}\times i\right)= 45+\left(\frac{110}{240}\times 10\right)$$

$$=45+4.58=49.58 \text{ marks}$$

Therefore, the average marks of a student are 49.6 approx.

PROBLEM 1.13. *Compute the arithmetic mean from the following data.*

Salary in Rs.	*Frequency*
Below 250	*30*
250-270	*16*
270-300	*19*
300-310	*20*
310-320	*10*
320 and over	*5*

SOLUTION. Computation of the arithmetic mean.

Salary in Rupees (*m*)	Mid-values (*m.v.*)	Frequency (*f*)	Deviations from the as. av. (285) (*dx*)	Total Deviations (*fdx*)
Below 250	240	30	—45	−1350
250-270	260	16	−25	−400
270-300	285	19	0	0
300-310	305	20	+20	+400
310-320	315	10	+30	+300
320 & over	325	5	+40	+200
		n=100		$\Sigma fdx=-850$

Applying the short-cut method.

Arithmetic mean or $a = x + \left(\frac{\Sigma fdx}{n}\right)$

We get, $a = 285 + \frac{-850}{100} = 285 — 8.5 = 276.50$ Rupees

Thus the arithmetic mean of the salaries of 100 men is Rs. 276.50

PROBLEM 1.14. *Find the average wage of a labourer from the following table :*

Wage in Rs. above.	*No. of labourers.*
300	*650*
310	*500*
320	*425*
330	*375*
340	*300*
350	*275*
360	*250*
370	*100*

SOLUTION. This is a cumulative series and in order to find out the average, it should be turned into a simple one.

Computation of average wage of 650 labourers

Wages in Rs. (*m*)	No. of labourers (*f*)	Mid-Points of the classes	Deviations from the as. av. (345)	Step Devia-tions (*dx*)	Total Devia-tions (*fdx*)
Exceed-ing Not Exceeding					
300-310	150	305	−40	−4	−600
310-320	75	315	−30	−3	−225
320-330	50	325	−20	−2	−100
330-340	75	335	−10	−1	− 75
340-350	25	345	0	0	0
350-360	25	355	+10	+1	+ 25
360-370	150	365	+20	+2	+300
370 and above	100	375	+30	+3	+300
	n=650				$\Sigma fdx=$ −375

Arithmetic average or $a = x = \left(\frac{\Sigma fdx}{n} \times i\right)$

$$= 345 + \left(\frac{-375}{650} \times 10\right) = 345 - 5.77 = 339.23 \text{ Rupees}$$

Therefore, the average wage of a labourer is Rs. 339.23.

PROBLEM 1.15. *Make a frequency table having grades of wages with class intervals of one rupee each from the following data of daily wages received by 30 labourers in a certain factory and then compute the average daily wages paid to a labourer.*

Daily wages in rupees.

8.25, 4.50, 6.50, 7.25, 5.50, 10.00, 9.75, 5.00, 7.75, 6.00,

11.00, 9.25, 8.50, 7.50, 6.25, 5.75, 4.00, 8.75, 9.50, 6.75,

4.75, 10.50, 6.25, 8.00, 7.25, 9.00, 7.50, 5.25, 7.00, 8.50

SOLUTION. Computation of the average daily wages of 30 labourers.

Wage in rupees (*m*)	No. of labourers (*f*)	Mid-values of the wage groups (*m.v.*)	Wages multiplied by the No. of labourers (*m. f.*)
4—5	3	4.50	13.50
5—6	4	5.50	22.00
6—7	5	6.50	32.50
7—8	6	7.50	45.00
8—9	5	8.50	42.50
9—10	4	9.50	38.00
10—11	3	10.50	31.50
	$n = 30$		$\Sigma mf = 225.00$

Substituting the above data in the formula.

Arithmetic Mean or $a = \frac{\Sigma mf}{n}$

We get, $a = \frac{225.00}{30} = 7.50$ rupees

Thus the average daily wages paid to labourer is rupees 7.50

PROBLEM 1.16. *Calculate arithmetic mean from the following data.*

Temp. '°C'		*No. of days*
−40 to	*−30*	*10*
−30	*—20*	*28*
−20	*—10*	*30*
—10	*0*	*42*
0	*10*	*65*
10	*20*	*180*
20	*30*	*10*

SOLUTION : Computation of arithmetic mean.

Temp. '°C' (*m*)	No. of days (*f*)	Mid points of the Classes (*m.v.*)	Deviations from the as. av. (5)	Step deviations (*dx*)	Total deviations (*fdx*)
−40 to −30	10	−35	−40	−4	− 40
−30 to −20	28	−25	−30	−3	− 84
−20 to −10	30	−15	−20	−2	− 60
−10 to 0	42	− 5	−10	−1	− 42
0 to 10	65	+ 5	0	0	0
10 to 20	180	+15	+10	+1	+180
20 to 30	10	+25	+20	+2	+ 20
	n=365				Σfdx= −26

Applying the short cut method.

$$\text{Arithmetic mean or } a=x+\left(\frac{\Sigma fdx}{n}\right)\times i$$

$$=5+\left(\frac{-26\times 10}{365}\right)$$

$$=5-\ 71$$

$$=4{\cdot}29\ ^\circ\text{C}.$$

PROBLEM 1.17. *The following table gives the monthly average of automobile production in Japan for the years 1968-1974 (unit 1,000 cars).*

Year	*Production*
1968	*358.4*
1969	*283.4*
1970	*363.2*
1971	*446.3*
1972	*279.7*
1973	*199.1*
1974	*114.2*

Calculate the average per cent of change per year.

SOLUTION. Calculation of average per cent of change per year.

Year	Production	Change from average 292.1 +signs ignored.
1968	358.4	66.3
1969	283.4	8.7
1970	363.2	71.1
1971	446.4	154.3
1972	279.7	12.4
1973	199.1	93.0
1974	114.2	177.9
$n=7$	2044.4	583.7

Mean Production $=\frac{2044.4}{7}$ (ooo cars.)=292.1 (ooo cars.)

Mean Change$=\frac{583.7}{7}=83.39$

Mean Percentage of Change $=\frac{83.4\times 100}{292.1}=28.6$

Note. (*a*) In this question ± signs have been ignored in the 3rd column because we are concerned only with a study of the amount of change, not its direction.

(*b*) Change has been calculated from the average of the whole period. In the absence of any other instruction it could also have been calculated from the 1st year (*i.e.*, 1968) or from each preceding year.

PROBLEM 1.18. *The following is the table of the ages of 30 adult persons.*

Digits (Division of class intervals)

Marks	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*Total*
20-29	*2*	*1*	*2*	*2*		*1*		*1*	*1*	*10*
30-39		*2*		*1*		*2*		*1*	*2*	*8*
40-49		*2*		*2*			*1*		*1*	*6*
50-59	*1*		*2*					*1*		*4*
60-69						*1*		*1*		*2*

Thus there are two persons of 23 years, one of 57 years and so on.

Find out the mean of the series.

(*a*) *by using only totals of class intervals.*

(*b*) *by using the entire data.*

SOLUTION. Calculation of the Mean

Mark	0	1	2	3	4	5	6	7	8	Total (f)	deviations dx from as. av. (45)	dx	fdx
20-29	40 2	21 1	44 2	46 2		25 1		27 1	28 1	231 10	−20	−2	−20
30-39		62 2		33 1		70 2		37 1	76 2	278 8	−10	−1	−8
40-49		82 2		86 2			46 1		48 1	262 6	0	0	0
50-59	50 1		104 2					57 1		211 4	10	1	4
60-69						65 1		67 1		132 2	20	2	4
										1114 $\Sigma f=30$			−20

(*a*) By using the total only.

$$\text{Mean} = x + \left(\frac{\Sigma\, fdx}{n} \times i\right) = 45 + \left(\frac{-20}{30} \times 10\right) = 45 - 6.7 = 38.3 \text{ years.}$$

(*b*) By using the entire data.

$$\text{Mean} = \frac{\Sigma mf}{\Sigma f} = \frac{1114}{30} = 37.1 \text{ years,}$$

Note. Figures on the top left corner of each square in the table indicate *mf*.

PROBLEM 1.19 *The following are the monthly salaries in rupees of 20 employees of a firm :*

130	*62*	*145*	*118*	*125*	*76*	*151*	*142*
110	*98*	*65*	*116*	*100*	*103*	*71*	*85*
80	*122*	*132*	*95*				

The firm gives bonuses of Rs 10, 15, 20, 25, and 30 for individuals in the respective salary groups : exceeding Rs, 60 but not exceeding Rs. 80 exceeding Rs. 80 but not exceeding Rs. 100 and so on upto exceeding Rs. 140 but not exceeding Rs. 160. Find the average bonus paid per employee.

SOLUTION : Calculation of Average Bonus

Class intervals	Alternative method of writing class intervals (Rs.)	Tally Bars	Frequency (f)	Rate of Bonus (m)	Total Bonus (mf)
Rs.	Rs.				
60 and not exceeding	80 60- 80	IIII	4	10	40
80 ,, ,, ,,	100 80-100	IIII	4	15	60
100 ,, ,, ,,	120 100-120	IIIII	5	20	100
120 ,, ,, ,,	140 120-140	IIII	4	25	100
140 ,, ,, ,,	160 140-160	III	3	30	90
			$n=20$		$\Sigma mf=390$

Total Bonus paid is Rs. 390

$$\text{Average Bonus per employee} = \frac{\Sigma mf}{n} = \frac{390}{20} = \text{Rs. } 19.50$$

WEIGHTED MEAN

PROBLEM 1.20. *A candidate obtains the following percentages in an examination : Sanskrit 75 ; Mathematics 84 ; Economics 56 ; English 78 ; Politics 57 ; History 54 ; Geography 47. It is agreed to give double weight to marks in English, Mathematics and Sanskrit. What is the weighted and unweighted mean ?*

SOLUTION. Computation of weighted and unweighted mean

Subjects	Marks percent (m)	Weights (w)	Weighted Marks (wm)
Sanskrit	75	2	150
Mathematics	84	2	168
Economics	56	1	56
English	78	2	156
Politics	57	1	57
History	54	1	54
Geography	47	1	47
$n=7$	$\Sigma m=451$	$\Sigma(w)=10$	$\Sigma wm=688$

$$\text{Weighted mean} : a = \frac{\Sigma(wm)}{\Sigma w} = \frac{688}{10} = 68.8 \text{ marks}$$

$$\text{Unweighted mean} : a = \frac{\Sigma m}{n} = \frac{451}{7} = 64.43 \text{ marks}$$

PROBLEM 1.21. *Explain what is meant by weighted average, and discuss the effect of weighting.*

Calculate (i) the unweighted mean of the prices in column III and (ii) the mean obtained by weighting each price by the quantity consumed.

I *Articles of food*	*II* *Quantity consumed*	*III* *Price in Rupees per kg.*
Flour	*500 kg.*	*1.25*
Ghee	*200 kg.*	*20.00*
Sugar	*30 kg.*	*4.50*
Potato	*15 kg.*	*0.50*
Oil	*40 kg.*	*5.50*

SOLUTION. Computation of the unweighted and the weighted average price of the food articles.

Articles of food	Price in Rupees per kg. (m)	Quantity consumed in kilograms (w)	Price multiplied by weights (wm)
Flour	1.25	500	625.00
Ghee	20.00	200	4000.00
Sugar	4.50	30	135.00
Potato	0.50	15	7.50
Oil	5.50	40	220.00
$n=5$	$\Sigma m=31.75$	$\Sigma w=785$	$\Sigma wm=4987.50$

Weighted Average

$$a' = \frac{\Sigma(wm)}{\Sigma w} = \frac{4987.50}{785} = 6.35 \text{ rupees}$$

Unweighted Mean

$$a = \frac{\Sigma m}{n} = \frac{31.75}{5} = 6.35 \text{ rupees}$$

Thus the unweighted mean of the prices is Rs. 6.35. Whereas the weighted mean of the prices is Rs. 6.35 approx.

PROBLEM 1.22. *From the following table calculate the average price of 500 gm. of tea, also the weighted average price :*

Price per 500 gm. *Rs. P.*	*Packets sold*
10.00	*200*
10.38	*275*
10.62	*400*
10.75	*150*
11.00	*100*
11.25	*75*
11.50	*50*

SOLUTION. Calculation of the weighted and the unweighted average price of 500 gms. of tea.

Price in Rs. per 500 gm (m)	Packets sold (w)	Deviations from assumed weighted average (10.75) (dx')	Total Deviations wdx
10.00	200	−.75	−150.00
10.38	275	−.37	−101.75
10.62	400	−.13	− 52.00
10.75	150	0	0
11.00	100	+.25	+25.00
11.25	75	+.50	+37.50
11.50	50	+.75	+37.50
$\Sigma m=75.50$	$\Sigma w=1250$		$\Sigma wdx'=-203.75$

Substituting the above data in the formula,

$$a' = x' + \frac{\Sigma(wdx')}{\Sigma(w)} :$$

Where a' stands for the weighted average x' for assumed weighte average ; w, for weight and dx' for deviation from assumed weighted average.

$$\text{We get } a' = 10.75 + \frac{-203.75}{1250} = 10.75 - .163 = 10.587 \text{ Rs.}$$

Thus the weighted average price of 500 gms, of tea is 10.587 Rs. whereas, the unweighted average price of the same is Rs. 10.79

as

$$a' = \frac{\Sigma m}{n} = \frac{75.50}{7} = 10.79 \text{ Rs.}$$

PROBLEM 1.23. *Compare the relative advantages and disadvantages of t Arithmetic Mean, the Median, and the Mode.*

The following table gives the results of certain examinations of three unive sities in the year 1974. Which is the best university? Give reasons for yo answer.

Percentage results in the University

University Examination	*A*	*B*	*C*
1. M.A.	*80*	*75*	*70*
2. M.Sc.	*70*	*70*	*60*
3. B.A.	*65*	*80*	*70*
4. B.Sc.	*60*	*70*	*80*
5. B.Com.	*75*	*65*	*75*

SOLUTION. It is always better to calculate a weighted average of percentages rather than a simple average. Therefore, in solving this questi weights have been assigned to the various percentages and the weights are proportion to the number of students in various classes in different universiti The number of students and consequently the weights also are imaginary.

Calculation of the weighted average of the percentage results in differe universities.

University Examination	University A %Re-sults	No. of students		University B %Re-sults	No. of students		University C %Re-sults	No. of students	
	m	w	wm	m	w	wm	m	w	w
M. A	80	20	1,600	75	30	2,250	70	50	3,50
M. Sc.	70	15	1,050	70	20	1,400	60	50	3,00
B. A.	65	100	6,500	80	200	16,000	70	300	21,00
B. Sc.	60	50	3,000	70	100	7,000	80	100	8,00
B. Com.	75	25	1,875	65	50	3,250	75	100	7,50
Total	350	210	14,025	360	400	29.900	355	600	43,0

$$\begin{array}{lccc} & A & B & C \\ \text{Simple average of percentages are :} & \frac{350}{5} & \frac{360}{5} & \frac{355}{5} \\ & = 70\% & 72\% & 71\% \\ \text{Weighted average of percentages are:} & \frac{14{,}025}{210} & \frac{29{,}900}{400} & \frac{43{,}300}{600} \\ & = 66.79\%, & 74.75\%, & 71.66\% \end{array}$$

Weighted average results show that B university is the best of all. As chance would have it the simple average also gives the same result. But it is a mere coincidence.

PROBLEM 1.24. *The following table indicates the increase in cost of living over July 1972 for a working class family as at 1st January 1975 and the weights assigned to various groups.*

Group	*Percentage increase*	*Weights*
Food	*29*	*7.5*
Rent	*54*	*2.0*
Clothing	*97.5*	*1.5*
Fuel and light	*75*	*1.0*
Other items	*75*	*0.5*

Find out the weighted average of the increase in cost of living.

SOLUTION. Computation of the weighted average of the increase in cost of living.

Group	Percentage increase (*m*)	Weights (*w*)	(*wm*)
ood	29.0	7.5	217.50
ent	54.0	2.0	108.00
lothing	97.5	1.5	146.25
uel and light	75.0	1.0	75.00
ther items	75.0	0.5	37.50
		$\Sigma w=12.5$	$\Sigma wm=584.25$

$$\text{Weighted average: } a'=\frac{\Sigma wm}{\Sigma w}=\frac{584.25}{12.5}=46.74\%$$

The weighted average of the increase in cost of living is thus 46.74%.

PROBLEM 1.25. *Compute the General and Standardized Death Rate of the ocal population and compare them with the General Death Rate of the standard opulation.*

ge Group	*Standard Population*		*Local Population*	
Years	*Population*	*Deaths*	*Population*	*Deaths*
0-10	*600*	*18*	*400*	*16*
10-20	*1,000*	*5*	*1,500*	*6*
20-60	*3,000*	*24*	*2,400*	*24*
60-100	*400*	*20*	*700*	*21*
Total	*5,000*	*67*	*5,000*	*67*

SOLUTION. Calculation of General and Standardized death rates.

ge Group	Standard Population			Local Population		
Year	Population	Deaths	Death rate per 1000	Population	Deaths	Death rate per 1000.
0—10	600	18	30	400	16	40
10—20	1,000	5	5	1,500	6	4
20—60	3,000	24	8	2,400	24	10
60—100	400	20	50	700	21	30
Total	5,000	67	13.4	5,000	67	13.4

General Death Rate of Standard Population

$$=\frac{(600\times 30)+(1{,}000\times 5)+(3{,}000\times 8)+(400\times 50)}{600+1{,}000+3{,}000+400}=13.4$$

General Death Rate of Local Population

$$=\frac{(400\times 40)+(1{,}500\times 4)+(2{,}400\times 10)+(700\times 30)}{400+1{,}500+2{,}400+700}=13.4$$

Standardized Death Rate of Local Population

$$=\frac{(600\times 40)+(1{,}000\times 4)+(3{,}000\times 10)+(400\times 30)}{600+1{,}000+3{,}000+400}=14$$

PROBLEM 1.26. *You are given the following statistics of population a unemployment in*

(*a*) *Your country as a whole for a standardized age-distribution;*

(*b*) *the local administrative area in which you live.*

Calculate (*i*) *the standardized unemployment rate in the country as a whole, (* *the standardized rate of unemployment in the local area and* (*iii*) *the crude rate* *unemployment in the local area.*

	(*Age years*)				
	15—30	*30—45*	*45—60*	*60—*	*Total*
Standard Population :	*250*	*350*	*300*	*100*	*1,00*
Age, constitution, unemployment rate, per cent—	*5*	*8*	*12*	*15*	—
Local Population :	*300*	*300*	*350*	*50*	*1,00*
Age, constitution, unemployment rate, per cent—	*4*	*9*	*12*	*20*	—

SOLUTION. Calculation of standardized and crude unemployment rates.

	Standard Population		Local Population	
Age	No. of people	Unemployment rate per cent	No. of people	Unemployment rate per cen
15-30	250	5	300	4
30-45	350	8	300	9
45-60	300	12	350	12
60-75	100	15	50	20
Total	1,000		1,000	

(*a*) Standardized unemployment rate for the country as a whole—

$$\frac{(250\times 5)+(350\times 8)+(300\times 12)+(100\times 15)}{1{,}000}=\frac{9{,}150}{1{,}000}=9.15$$

(*b*) Standardized rate of unemployment in local area—

$$\frac{(250\times 4)+(350\times 9)+(300\times 12)+(100\times 20)}{1{,}000}=\frac{9{,}750}{1{,}000}=9.75$$

(*c*) Crude rate of unemployment in the local area—

$$=\frac{(300\times 4)+(300\times 9)+(350\times 12)+(50\times 29)}{1{,}000}=\frac{9{,}100}{1{,}000}=9.1$$

Moving Average :

PROBLEM 1.27. *Compute five yearly and four yearly moving average from the following:—*

Years	*1950*	*1951*	*1952*	*1953*	*1954*	*1955*	*1956*	*1957*	*1958*	*1959*	
Value Rs.	*365*	*360*	*355*	*330*	*300*	*330*	*340*	*290*	*280*	*250*	
Years	*1960*	*1961*	*1962*	*1963*	*1964*	*1965*	*1966*	*1967*	*1968*	*1969*	*1970*
Value Rs.	*235*	*255*	*250*	*245*	*225*	*210*	*200*	*230*	*225*	*200*	*195*

SOLUTION

	Five yearly moving Average			Four yearly moving average		
Year	Value	Total	Moving Average	Total of Four	Total paired	Moving Average
1950	365	—	—		—	—
1951	360	--	—	1410	—	—
1952	355	1710	342	1345	2755	344.4
1953	330	1675	335	1315	2660	332.5
1954	300	1655	331	1300	2615	326.9
1955	330	1590	318	1260	2660	320.5
1956	340	1540	308	1240	2500	312.5
1957	290	1490	298	1160	2400	300.0
1958	280	1395	279	1055	2215	216.9
1959	250	1310	262	1020	2075	259.4
1960	235	1270	254	990	2010	251.25
1961	255	1235	247	985	1975	246.9
1962	250	1210	242	975	1960	245.0
1963	245	1185	237	930	1905	238.1
1964	225	1130	226	880	1810	226.25
1965	210	1110	222	865	1745	218.1
1966	200	1090	218	865	1730	216.25
1967	230	1060	213	855	1720	215.0
1968	225	1055	210	850	1705	213.1
1969	200	—	—			
1970	195	—	—			

MEDIAN

PROBLEM 1.28. *Find out the median of the following items :—*

5, 7, 9, 12, 10, 8, 7, 15, 21,

SOLUTION. Items arranged in ascending order of magnitude

Serial number	Size of items
1	5
2	7
3	7
4	8
5	9
6	10
7	12
8	15
9	21

If M represents the median, and n the number of items.

$$M = \text{Size of } \frac{n+1}{2} \text{ items} = \text{Size of } \frac{9+1}{2} \text{ items}$$

$$= \text{Size of 5th item} = 9.$$

S.C.E.R.T., West Bengal
Date 15.3.78 ... 311

S.C.E.R.T., West Bengal Library
25/3, B. C. Ro
Calcutta

PROBLEM 1.29. *The following table gives the Economic Advisers Ind Numbers of wholesale prices (1959=100)*

Year	*Index numbers*
1967	*297.4*
1968	*367.1*
1969	*381.1*
1970	*400.7*
1971	*439.3*
1972	*386.9*

Find the median.

SOLUTION. Index Number arranged in ascending order of magnitude

Serial number	Size of items
1	297.4
2	367.1
3	381.1
4	386.9
5	400.7
6	439.3

If M represents the median, and n the number of items.

$$M = \text{Size of } \frac{n+1}{2} \text{ items} = \text{Size of } \frac{6+1}{2} \text{ items}$$

$$= \text{Size of 3.5th item} = \frac{(381.1+386.9)}{2} = \frac{768.0}{2} = 384.0$$

PROBLEM 1.30. *The following table gives the marks obtained by a batc of 30 B. Com. students in a class-test in statistics.*

(*Maximum Marks 100*)

Roll No.	*Marks obtained*	*Roll No.*	*Marks obtained*
1	*33*	*16*	*24*
2	*32*	*17*	*33*
3	*55*	*18*	*42*
4	*47*	*19*	*38*
5	*21*	*20*	*45*
6	*50*	*21*	*26*
7	*27*	*22*	*33*
8	*12*	*23*	*44*
9	*68*	*24*	*48*
10	*49*	*25*	*52*
11	*40*	*26*	*30*
12	*17*	*27*	*58*
13	*44*	*28*	*37*
14	*48*	*29*	*38*
15	*62*	*30*	*35*

Find the value of the Median and the Quartiles.

SOLUTION. Marks obtained by 30 students arranged in ascending order of magnitude

Serial No.	Marks	Serial No.	Marks	Serial No.	Marks
1	12	11	33	21	47
2	17	12	35	22	48
3	21	13	37	23	48
4	24	14	38	24	49
5	26	15	38	25	50
6	27	16	40	26	52
7	30	17	42	27	55
8	32	18	44	28	58
9	33	19	44	29	62
10	33	20	45	30	68

If M represents the median, and n the number of items

M=Size of $\frac{n+1}{2}$ items

=Size of $\frac{30+1}{2}$ items

=Size of 15.5th item.

$= \frac{\text{Size of 15th item+Size of the 16th item.}}{2}$

$= \frac{38+40}{2} = 39$ marks

If Q_1 represents the First Quartile, and n the number of items :

Q_1=Size of $\frac{n+1}{4}$ items

=Size of $\frac{30+1}{4}$ items

=Size of 7.75th item.

=Size of 7th item$+\frac{3}{4}$ (Size of 8th item—Size of 7th item)

$=30+\frac{3}{4}(32-30)$

=31.5 marks

If Third Quartile is represented by Q_3

Q_3=Size of $\frac{3(n+1)}{4}$ items

=Size of $\frac{3(30+1)}{4}$ items

=Size of 23.25th item.

=Size of 23rd item$+\frac{1}{4}$(Size of 24th item—Size of 23rd item)

$=48+\frac{1}{4}(49-48)$

=48.25 marks.

PROBLEM 1.31 *Find the median size of the shoes from figures given below :—*

Size of Shoes	*Frequency*
4.5	*1*
5	*2*
5.5	*4*
6	*5*
6.5	*15*
7	*30*
7.5	*60*
8	*95*
8.5	*82*
9	*75*
9.5	*44*
10	*25*
10.5	*15*
11	*4*

Also calculate the Quartiles, 7th Decile, 46th Percentile, 3rd Quintile and 5th Octile.

SOLUTION. Calculation of the median size of the shoes :—

Size of shoes	Frequency	Cumulative Frequency
4.5	1	1
5	2	3
5.5	4	7
6	5	12
6.5	15	27
7	30	57
7.5	60	117
8	95	212
8.5	82	294
9	75	369
9.5	44	413
10	25	438
10.5	15	453
11	4	457

Median $=$size of $\frac{n+1}{2}$ pairs ; where n equals the total frequency

$=$Size of $\frac{457+1}{2}$ or 229th pair

$=8.5$

Lower Quartile $=$Size of $\frac{n+1}{4}$ pairs

$=$size of $\frac{457+1}{4}$ or 114.5th pair

$=7.5$

Upper Quartile $=$Size of $\frac{3(n+1)}{4}$ pairs

$=$size of $\frac{3(457+1)}{4}$ or 343.5th pair

$=9$

7th Decile $=$size of $\frac{7(n+1)}{10}$ pairs

$=$size of $\frac{7(457+1)}{10}$ or 320.6th pair

$=9$

46th Percentile $=$size of $\frac{46(n+1)}{100}$ pairs

$=$size of $\frac{46(457+1)}{100}$ or 210.68th pair

$=8$

3rd Quintile $=$size of $\frac{3\,(n+1)}{5}$ pairs

size of $\frac{3(457+1)}{5}$ or 274.8th pair

$=8{\cdot}5$

5th Octile $=$size of $\frac{5\,(n+1)}{8}$ pairs

$=$size of $\frac{5(457+1)}{8}$ or 286.25th pair

$=8.5$

PROBLEM 1.32. *The table below shows the age distribution of married females according to sample census of 1971 in the Gujrat State.*

Age	*Number of married females*
0-5	*3*
5-10	*31*
10-15	*410*
15-20	*1,809*
20-25	*2,446*
25-30	*2,223*
30-35	*1,723*
35-40	*1,292*
40-45	*963*
45-50	*762*
50-55	*531*
55-60	*317*
60-65	*156*
65-70	*59*
70-75	*37*

Calculate the median age of married females and also the two quartiles.

SOLUTION. Calculation of the median and quartile ages of married females.

Age	Number of married females	Cumulative Frequency
0- 5	3	3
5-10	31	34
10-15	410	444
15-20	1,809	2,253
20-25	2,446	4,699
25-30	2,223	6,922
30-35	1,723	8,645
35-40	1,292	9,937
40-45	963	10,900
45-50	762	11,662
50-55	531	12,193
55-60	317	12,510
60-65	156	12,666
65-70	59	12,725
70-74	37	12,762
Total	12,762	

The median age of married females

=the age of the $\frac{n}{2}$ females, where n equals the total frequency

age of the $\frac{12762}{2}$ *i.e.*, 6381 married females who lie in the (25-30) age group. Applying the formula of interpolation,

$M = l_1 + \frac{l_2 - l_1}{f_1}(m-c)$; where, M represents the median, l_1 and l_2 the lower and the upper limits of the group in which median is situated; f_1 the frequency of median class, m, the number of middle item or $n/2$ items and c, the cumulative frequency of the group lower than the one in which median is situated.

$$M=25+\frac{30-25}{2223}(6381-4699)$$

$=28.8$ years, approx.

The lower quartile age of married females

=the age of the $n/2$ *i. e.*, 3190.5 married females who lie in the 20-25 age group ;

By interpolation

$$Q_1=l_1+\frac{l_2-l_1}{f_1}(q_1-c) \;;$$

where, Q_1 represents lower quartile, l_1 and l_2, the lower and the upper limits of the group in which lower quartile is situated, f_1 the frequency of lower quartile class; q_1 the number of $n/_4$ items; and c, the cumulative frequency of the group lower than the one in which the lower quartile is situated.

$$=20+\frac{25-20}{2246}(3190.5-2253)=20+2.09$$

$=22.09$ or 22.1 yrs. approx.

The upper quartile age of married females

=the age of the $3\,(n)/4$ *i. e.*, 9571.5 married females who are situated in the 35-40 age group ;

By interpolation,

$$Q_3=l_1+\frac{l_2-l_1}{f_1}(q_3-c) \;;$$

where Q_3 stands for upper quartile ; l_1 and l_2 for the lower and the upper limits of the group in which upper quartile is situated; f_1 for the frequency of upper quartile class; q_3 for the number of $3\,(n)/4$ items: and c for the cumulative frequency of the group lower than the one in which the upper quartile is situated.

$$=35+\frac{40-35}{1292}(9571.5-8645)$$

$=38.6$ yrs. approx.

PROBLEM 1.33. *Calculate the values of the median and the two quartiles for the following:---*

Limits for percentage recovery of sugar on cane.	*Factories in India (1974-75)*
8.0-8.2	*2*
8.2-8.4	*5*
8.4-8.6	*4*
8.6-8.8	*11*
8.8-9.0	*11*
9.0-9.2	*11*
9.2-9.4	*13*
9.4-9.6	*10*
9.6-9.8	*7*
9.8-10.0	*6*
10.0-10.2	*3*
10.2-10.4	*1*
10.4-10.6	*1*
	85

SOLUTION. Calculation of the median and quartiles percentage recovery of sugar.

Percentage recovery of sugar on cane	No. of factories	Cumulative Frequency
8.0- 8.2	2	2
8.2- 8.4	5	7
8.4- 8.6	4	11
8.6- 8.8	11	22
8.8- 9.0	11	33.
9.0- 9.2	11	44
9.2- 9.4	13	57
9.4- 9.6	10	67
9.6- 9.8	7	74
9.8-10.0	6	80
10.0-10.2	3	83
10.2-10.4	1	84
10.4-10.6	1	85
	$n=85$	

The Median percentage recovery of sugar on cane is 9.17, as
M=the percentage recovery of sugar on cane of the 85/2 *i.e.*, 42.5th factory.

=(9.0—9.2) recovery group, and
By interpolation,

$$M=l_2+\frac{l_2-l_1}{f_1}(m-c)=9.0+\frac{9.2\text{-}9.0}{11}(42.5\text{-}33)=9.17 \text{ per cent.}$$

The lower quartile percentage recovery of sugar on cane is 8.8 as
Q_1=the percentage recovery of 85/4 *i.e.*, 21.25th factory.
=(8.6-8.8) recovery group; and by interpolation

$$Q_1=l_1+\frac{l_2-l_1}{f_1}(q_1-c)=8.6+\frac{8.8-8.6}{11}(21.25-11)=8.78\%$$

The upper quartile percentage of recovery is 9.54, as

Q_3=the percentage recovery of $\frac{3(85)}{4}$ *i. e.*, 63.75th factory.

=(9.4−9.6) recovery group ; and by interpolation :

$$Q_3=l_1+\frac{l_2-l_1}{f_1}(q_3-c)=9.4+\frac{9.6-9.4}{10}(63{\cdot}75-57)$$

=9.53 per cent.

PROBLEM 1.34. *Calculate the mean and median for the following distribution.*

Weight of boys in a certain class	*Number*
100-104	*4*
105-109	*14*
110-114	*60*
115-119	*138*
120-124	*206*
125-129	*298*
130-134	*380*
135-139	*450*
140-144	*500*
145-149	*430*
150-154	*260*
155-159	*128*
160-164	*66*
165-169	*28*
170-174	*12*
	2974

SOLUTION. Calculation of the Mean and the Median Weight of 2974 boys.

Weight (lbs.) (*m*)	Number of boys (*f*)	Mid-values (*m.v.*)	Deviations from the assumed average (137) *dx*	Total Deviations *fdx*	Cumulative Frequency	Lower limits of the classes
100-104	4	102	+35	+ 140	4	99.5
105-109	14	107	+30	+ 420	18	134.5
110-114	60	112	+25	+1,500	78	109.5
115-119	138	117	+20	+2,760	216	114.5
120-124	206	122	+15	+3,090	422	119.5
125-129	298	127	+10	+2,980	720	124.5
130-134	380	132	+ 5	+1,900	1,100	129.5
135-139	450	137	0	0	1,550	134.5
140-144	500	142	+ 5	+2,500	2,050	139.5
145-149	430	147	+10	+4,300	2,480	144.5
150-154	260	152	+15	+3,900	2,740	149.5
155-159	128	157	+20	+2,560	2,868	154.5
160-164	66	162	+25	+1,650	2,934	159.5
165-169	28	167	+30	+ 840	2,962	164.5
170-174	12	172	+35	+ 420	2,974	169.5
	n=2974			Σfdx+3,380		

$$\text{Mean weight of boys} = x + \frac{\Sigma fdx}{n} = 137 + \frac{3380}{2974} = 138.14$$

Median weight of the boys = the weight of the $\frac{n}{2}$ *i.e.,* 1487th boy

who lies in the 134.5-139.5 weight group. By interpolation

$$M = 134.5 + \frac{139.5 - 134.5}{450}(1487 - 1100) = 138.8$$

PROBLEM 1.35. *The following table gives the distribution of the male and female population of a certain area in India. By finding the mean age, the median age, and the upper and lower quartile ages make comments on the age distribution of the two sexes in the area :*

Age group	*Male*	*Female*
0- 9	*2,756*	*2,787*
10-19	*2,124*	*2,032*
20-29	*1,677*	*1,724*
30-39	*1,481*	*1,485*
40-49	*1,021*	*1,022*
50-59	*610*	*579*
60-69	*245*	*269*
70-79	*67*	*78*
80-89	*16*	*20*
90-99	*3*	*4*
Total	*10,000*	*10,000*

SOLUTION. Determination of median, lower and upper quartile ages.

Age-group	Males		Females		Upper limit of age
	Frequency	Cumulative Frequency	Frequency	Cumulative Frequency	
0- 9	2,756	2,756	2,787	2,787	9.5
10-19	2,124	4,880	2,032	4,819	19.5
20-29	1,677	6,557	1,724	6,543	29.5
30-39	1,481	8,038	1,485	8,028	39.5
40-49	1,021	9,059	1,022	9,050	49.5
50-59	610	9,669	579	9,629	59.5
60-69	245	9,914	269	9,898	69.5
70-79	67	9,981	78	9,976	79.5
80-89	16	9,997	20	9,996	89.5
90-99	3	10,000	4	10,000	99.5

Required age	Male	Female
Median	Age of 5,000th male =(19.5-29.5) age group ; By interpolation $=19.5+\frac{29.5-19.5}{1677}(120)$ =20.21 yrs.	=Age of 5000th female =(19.5-29.5) age group ; By interpolation $=19.5+\frac{29.5-19.5}{1724}(181)$ =20.55 yrs.

Required age	Male	Female
Lower quartile	=Age of 2500th male =(0—9.5) age group ; By interpolation $=0+\frac{9.5}{2756}(2500)$ yrs. =8.62 yrs.	=Age of 2500th female =(0—9.5)age group; By interpolation $=0+\frac{9.5}{2787}(2500)$ yrs. =8.52 yrs.
Upper Quartile age	=Age of 7500th male =(29.5—39.5) age group ; By interpolation $=29.5+\frac{10}{1481}(943)$ =35.87 yrs.	=Agė of 7500th female, =(29.5—39.5) age group By interpolation $=29.5+\frac{10}{1485}(957)$ =35.98 yrs.

PROBLEM 1.36. *Determine the quartiles and the median for following table :—*

Income	*No. of persons*
Below Rs. 330	*69*
Rs. 330 and below Rs. 340	*167*
Rs. 340 and below Rs. 350	*207*
Rs. 350 and below Rs. 360	*65*
Rs. 360 and below Rs. 370	*58*
Rs. 370 and below Rs. 380	*27*
Rs. 380 and over	*10*
Total	*603*

SOLUTION. Calculation of the median and quartile income.

Income	No. of persons	Cumulative Frequency
Below Rs. 330	69	69
Rs. 330 and below Rs. 340	167	236
Rs. 340 and below Rs. 350	207	443
Rs. 350 and below Rs. 360	65	508
Rs. 360 and below Rs. 370	58	566
Rs. 370 and below Rs. 380	27	593
Rs. 380 and over	10	603
	$n=603$	

The median income is Rs. 343·16, as

M=the income of $n/2$ *i.e.*, 301.5th person, and therefore, it lies in the 340-350 income group ;

By interpolation.

$$M=340+\frac{350\text{-}340}{207}(301.5-236)=343{\cdot}16 \text{ rupees}$$

The lower quartile income is Rs. 334·9. as.

Q_1 =the income of 150·75 person and, therefore, lies in the 330—340 income group ;

By interpolation,

$$Q_1=330+\frac{340-330}{167}(150{\cdot}75-69)=334{\cdot}9 \text{ rupees}$$

The upper quartile income is Rs. 351·42 as

Q_2=the income of 452·25 person and, therefore, lies in the 350—360 income group

By interpolation :

$$Q_3=350+\frac{360-350}{65}(452.25-443)=351.42 \text{ rupees}$$

PROBLEM 1.37. *The following table classifies the she-buffaloes of India in 1970 according to the yield of milk per day. Calculate from the data the mean and the median yield of milk per she-buffalo.*

Yield per day in kgs.	*No. of she-buffaloes in thousands*
Up to 1	*114*
Above 1 to 2	*2005*
,, *2 to 3*	*7706*
,, *3 to 4*	*4590*
,, *4 to 5*	*2080*
,, *5 to 6*	*240*
,, *6 to 7*	*3580*
Total	*20315*

SOLUTION. Calculation of the mean and the median yields of milk per she-buffalo.

Mean yield of milk per she-buffalo

$$=3{\cdot}5+\frac{1,242}{20,315}=3{\cdot}5+{\cdot}06\text{kgs.}=3{\cdot}56 \text{ kgs.}$$

Yield per day in kgs. (*m*)	No. of she buffaloes in thousands (*f*)	Mid-values (*mv*)	Deviations from the assumed average(3.5) *dx*	Total Deviation *fdx*	Cumulative Frequency *cf*
Above-upto					
0-1	114	.5	—3	— 342	114
1-2	2,005	1.5	—2	— 4,010	2,119
2-3	7,706	2.5	—1	— 7,706	9,825
3-4	4,590	3.5	0	0	!4,415
4-5	2,080	4.5	+1	+ 2,080	16,495
5-6	240	5.5	+2	+ 480	16,735
6-7	3,580	6.5	+3	+10,740	20,315
	n=20,315			Σfdx=+1242	

Median yield of milk per she-buffalo

$=$yield of the $\frac{20,315}{2}$ *i.e.*, 1,0157·5th she-buffalo

$=$(3—4) yield group ; By interpolation :

$$M=3+\frac{4-3}{4590}(10157{\cdot}5-9825)=3.07 \text{ kgs.}$$

PROBLEM 1.38. *Find the median and the quartiles.*

Amount of wages	*Number of workers receiving such rate of wages*
Not exceeding 8 Rupees	*85*
Over 8 Rs. but not exceeding 10 Rs.	*65*
Over 10 Rs. but not exceeding 12 Rs.	*59*
Over 12 Rs. but not exceeding 14 Rs.	*50*

SOLUTION. Calculation of the median and the quartiles wages.

Amount of wages	Number of workers receiving such Rate of wages	Cumulative Frequency
Not exceeding 8 Rupees	85	85
Over 8 Rs. but not exceeding 10 Rs.	65	150
Over 10 Rs. but not exceeding 12 Rs.	59	209
Over 12 Rs. but not exceeding 14 Rs.	50	259
	n=259	

Median wages$=$the wage of the $\frac{259}{2}$ *i. e.*, 129·5th labourer

which lies in (over 8 but not exceeding 10 Rs.) wage group ; By interpolation.

$$=8+\frac{10-8}{65}(129{\cdot}5-85)=9.4 \text{ rupees.}$$

Lower quartile wage

$=$the wages of the 259/4 *i.e.*, 64.74th labourer, which lies in (not excee ding 8 rupees) wage group : By interpolation

$$=6+\frac{8-6}{85}(64{\cdot}75-0)=7.52 \text{ rupees}$$

Upper Quartile wages

$=$the wages of the 3(259)/4 *i.e.*, 194·25th labourer which lies in (ove 10 Rs. but not exceeding 12 Rs.) wage group ; By interpolation,

$$=10+\frac{12-10}{9}(194{\cdot}25-150)=11 \text{ rupees.}$$

PROBLEM. 1.39. *Amend the following table, and locate the median from the amended table. Also measure the magnitude of the median so located :*

Sizes	*Frequency*
10-15	*10*
16-17.5	*15*
17.5-20	*17*
22-30	*25*
30-35	*28*
35-40	*30*
45 & onwards	*40*

SOLUTION. Calculation of the median size.

Sizes	Frequency	Cumulative
10-20	42	42
20-30	25	67
30-40	58	125
40-50	40	165

M=the size of the $\frac{n}{2}$ item ;

=the size of the $\frac{165}{2}$ *i.e.*, 82.5th item ;

which lies in the 30—40 group ;

By interpolation we have

$$M=l_1+\frac{l_2-l_1}{f_1}(m-c)\ 30+\frac{40-30}{58}(82.5-67)=32.67$$

Thus the median size is 32.67

PROBLEM 1.40. *The following table gives the marks obtained by 65 student in Statistics in a certain examination :*

Examination marks	*Number of Students*
More than 70%	*7*
More than 60%	*18*
More than 50%	*40*
More than 40%	*40*
More than 30%	*63*
More than 20%	*65*

Calculate the median of the above series.

SOLUTION. Calculation of median percentage of marks.

Marks (m)	Number of students (f)	Cumulative frequency (e)
20-30%	2	2
30-40%	23	25
40-50%	0	25
50-60%	22	47
60-70%	11	58
Above 70%	7	65

M=Size of $n/2$ items

=Size of $65/2$ items

=Size of the 32.5th item which lies in 50-60 group.

$$=50+\frac{60\text{-}50}{22}(32.5-25)=50+\left(\frac{10}{22}\times7{\cdot}5\right)$$

=50+3.4=53.4% marks

Alternate method

Marks (m)	No. of students (f)	Cumulative frequency (cf)
Above 70%	7	7
60-70%	11	18
50-60%	22	40
40-50%	0	40
30-40%	23	63
20-30%	2	65

Median=Size of $\frac{n}{2}$ or 32.5th item, which lies in 50-60 group.

$=l_1+\frac{l_2-l_1}{f_1}(m-c)$, where c represents the difference between the total frequency and cumulative frequency of the median group. If items are arranged in descending order, instead of ascending order, this change is necessary.

$$=50+\frac{60-50}{22}[32.5-(65-40)]=50+\frac{10}{22}(32.5-65+40)$$

$$=50+\left(\frac{10\times 7.5}{22}\right)=50+3.4=53.4 \text{ percent marks.}$$

PROBLEM 1.41. *Find out the median of the following series :*

Wages *Rs.*	*No. of labourers*
260-270	*5*
250-260	*10*
240-250	*20*
230-240	*5*
220-230	*3*

SOLUTION.

Wages in Rs. (m)	No. of labourers (f)	Cumulative Frequency (cf)
260-270	5	5
250-260	10	15
240-250	20	35
230-240	5	40
220-230	3	43

Median=Value of 21.5th item which is in 40-50 group.

$=l_1+\frac{l_2-l_1}{f_1}(m-c)$, where 'c' represents the difference between the total frequency and the cumulative frequency of the median group.

$$=240+\frac{250-240}{20}\left[21.5-(43-35)\right]=240+\frac{10}{20}(21.5-43+35)$$

$$=240+\left(\frac{10}{20}\times 13.5\right)=240+6.75=246.75 \text{ rupees.}$$

PROBLEM 1.42. *Form a frequency distribution of the following data givin the index numbers of 60 commodities in a certain year and find the value of th Mean and the Median.*

76, 79, 81, 86, 86, 87, 89, 90, 91, 94, 95, 96, 96, 96, 97, 99, 99, 100, 100, 101, 101, 101, 102, 104, 104, 104, 105, 106, 106, 107, 108, 108, 109, 109, 109, 110, 110, 111, 112, 113, 113, 114, 114, 115, 116, 116, 116, 117, 117, 118, 119, 120, 121, 122, 123, 124, 125, 128, 129, 134.

SOLUTION. The items in the question are already arranged in ascending order and hence Median will be the value of the middle item or $\frac{n+1}{2}$ or $\frac{60+1}{2}$ or 30.5th item. M is, therefore, equal to $\frac{107+108}{2}=107.5$

$$\text{Mean or Arithmetic average}=\frac{\Sigma m}{n}=\frac{6378}{60}=106.3$$

In such a question it is sometimes more convenient (though less accurate) to convert the series into a continuous one and then to solve the problem, as is done below :—

Calculation of the Mean and the Median Index Numbers.

Index Numbers (*m*)	Frequency (*f*)	Mid-values (*m.v.*)	Deviations from the assumed average (100) *dx*	Total Deviation *dx*	Cumulative Frequency
75- 85	3	80	−20	−60	3
85- 95	7	90	−10	−70	10
95-105	16	100	0	0	26
105-115	17	110	+10	+170	43
115-125	13	120	+20	+260	56
125-135	4	130	+30	+120	60
	$n=60$			$\Sigma dx=+420$	

Mean of the Index Numbers

$$\text{Mean}=x+\frac{\Sigma dx}{n}=100+\frac{420}{60}=107$$

Median Index Number = the value of the $n/2$ *i.e.*, 30th item, which lies in the 105-115 groups. By interpolation

$$=105+\frac{115-105}{17}(30-26)=107.35$$

Thus the Mean and the Median Index Numbers according to this method are 107 and 107.35 respectively. It shall be noted that there is a significant difference between the value of the mean calculated by the two methods. The former value (106.3) is more accurate than the latter (107).

PROBLEM 1.43. *In a small branch of a bank in rural area, the following is the average deposit balance of current accounts during a month. Calculate median, 7th decile and 85th percentile.*

Deposit balance	*No. of deposits*
Less than Rs. 1000	*500*
Rs. 900	*498*
Rs. 800	*480*
Rs. 600	*475*
Rs. 550	*440*
Rs. 500	*374*
Rs. 400	*300*
Rs. 250	*125*
Rs. 100	*25*

SOLUTION. Calculation of median, 7th decile and 85th percentile.

Deposit balance (m) Rs.	No. of deposits (f)	Cumulative frequency (cf)
Upto 100	25	25
100-250	100	125
250-400	175	300
400-500	74	374
500-550	66	440
550-600	35	475
600-800	5	480
800-900	18	498
900-1000	2	500

Median=Size of $n/2$th items

$$=\text{Size of } \frac{500}{2} \text{ items}$$

=Size of 250th item which lies in 250-400 group

By interpolation

$$M=250+\frac{400-250}{175}(250-125)$$

$$=250+\frac{150}{175}\times 125$$

$$=250+107.1=357.1 \text{ rupees approx.}$$

$$D_7=\text{Size of } \frac{7(n)\text{th}}{10} \text{ item}$$

$$=\text{Size of } \frac{7(500)}{10} \text{ item}$$

=Size of 350th item which lies in 400—500 group

By Interpolation—

$$D_7=400+\frac{500-400}{74}(350-300)$$

$$=400+\frac{100}{74}(50)$$

$$=400+67.6=467.6 \text{ rupees approx.}$$

$$P_{85}=\text{Size of } \frac{85\,(n)\text{th}}{100} \text{ item}$$

$$=\text{Size of } \frac{85(500)\text{th}}{100} \text{ item}$$

=Size of 425th item which lies in 500—550 group

By interpolation—.

$$P_{85}=500+\frac{550-500}{66}(425-374)$$

$$=500+\frac{50}{66}(51)$$

$$=500+38.6=538.6 \text{ rupees approx.}$$

PROBLEM 1.44. *The following table gives the weekly wages in rupees in certain commercial organization :*

Weekly Wages Rs.	*Frequency*	*Weekly Wages Rs.*	*Frequency*
30-32	*2*	*40-42*	*62*
32-34	*9*	*42-44*	*39*
34-36	*25*	*44-46*	*20*
36-38	*30*	*46-48*	*11*
38-40	*49*	*48-50*	*3*

Calculate from the above data :

(*i*) *The median and the third quartile wages.*

(*ii*) *The number of wage earners receiving between Rs. 37 and 45 p week.*

SOLUTION. Calculation of median and quartile wages

Weekly Wages Rs.	Frequency	Cumulative Frequency
30-32	2	2
32-34	9	11
34-36	25	36
36-38	30	66
38-40	49	115
40-42	62	177
42-44	39	216
44-46	20	236
46-48	11	247
48-50	3	250
	$n=250$	

(*i*) The median wages are Rs. 40.32, as

M= the wages of 250/2 or 125th person, and, therefore, it lies in the 40 − 42 wages group.

By interpolation.

$$M=40+\frac{42-40}{62}(125-115)=40.32 \text{ rupees approx.}$$

The third quartile wages are Rs. 42.54, as

Q_3= the wages of 3(250/4) or 187.5th person and, therefore, lies in 42—44 wages group.

$$Q_3=42+\frac{44-42}{39}(187.5-177)=42.54 \text{ rupees approx.}$$

(*ii*) The number of wage earners between Rs. 37 and Rs. 45 per week are as under :

Rs.
Between 36—38=30 workers
∴ 37—38=15 ,,
38—40=49 ,,
40—42=62 ,,
42—44=39 ,,
44—45=10 ,,

Total Rs. 37—45=175 workers

PROBLEM 1.45. *The following table gives you the distribution of marks secured by some students in an examination :*

Marks (out of 100)	*No. of students*
0-20	*42*
21-30	*38*
31-40	*120*
41-50	*84*
51-60	*48*
61-70	*36*
71-80	*32*

Find : (*i*) *Median marks*

(*ii*) *The percentage of failures if the minimum for a pass is 35 marks.*

SOLUTION Determination of Median marks.

Marks (out of 100)	No. of students	Cumulative frequency	Upper limit of Marks
0-20	42	42	20.5
21-30	38	80	30.5
31-40	120	200	40.5
41-50	84	284	50.5
51-60	48	332	60.5
61-70	36	368	70.5
71-80	32	400	80.5

(*i*) The median is 40.5 marks as

M= the marks of $n/2$ *i.e.*, 200th student, and, therefore, it lies in the 0.5—40.5 marks group ;

By interpolation.

$$M=30.5+\frac{40.5-30.5}{120}(200-80)=40.5 \text{ marks.}$$

(*ii*) The student who secures 34 marks is failed in the exam. Therefore, 4th percentile will give the number of students who secured marks 34 and less.

$$px=l_1+\frac{l_2-l_1}{f}(p_x-c) \text{ where } px \text{ stands for any percentile.}$$

i.e., $34=30.5+\dfrac{40.5-30.5}{120}(x-80)$

or $34=30.5+\dfrac{10}{120}(x-80)$

or $3.5=\dfrac{10x-800}{120}$

or $10x=1220$

$\therefore\ x=\dfrac{1220}{10}=122\text{th student.}$

Thus, the percentage of failures $=\dfrac{122}{400}\times100=30.5\%$

MODE

PROBLEM 1.46. *Find out the mode of the following series :—*

Size of item	*Frequency*
1	*4*
2	*10*
3	*16*
4	*18*
5	*24*
6	*28*
7	*28*
8	*30*
9	*22*
10	*26*
11	*18*
12	*14*
13	*8*
14	*6*

SOLUTION.

Location of mode by grouping.

Size of item (m)	Frequency (f)					
	(1)	(2)	(3)	(4)	(5)	(6)
1	4	14		30		
2	10		26		44	
3	16	34				58
4	18		42	70		
5	24	52			**80**	
6	28		**56**			**86**
7	28	**58**		**80**		
8	**30**		52		78	
9	22	48				66
10	26		44	58		
11	18	32			40	
12	14		22			28
13	8	14				
14	6					

Analysis Table

Columns	Size of item containing maximum frequency				
(1)				8	
(2)			7	8	
(3)		6	7		
(4)			7	8	9
(5)	5	6	7		
(6)		6	7	8	
No. of time size occurs	1	3	5	4	1

The size 7 is, therefore, the mode, which is the most frequent size of the items.

PROBLEM 1.47. *Find out the Mode of the following series :*

Size	*Frequency*	*Size*	*Frequency*
5	*48*	*13*	*52*
6	*52*	*14*	*41*
7	*56*	*15*	*57*
8	*60*	*16*	*63*
9	*63*	*17*	*52*
10	*57*	*18*	*48*
11	*55*	*19*	*40*
12	*50*	—	—

SOLUTION. Location of Mode by grouping.

Size of item	Frequency (*f*)					
(*m*)	(1)	(2)	(3)	(4)	(5)	(6)
5	48	100		156		
6	52		18		168	
7	56	116				170
8	60		**123**	**180**		
9	63	**120**			**175**	
10	57		112			162
11	55	105		157		
12	50		102		143	
13	52	93				150
14	41		98	161		
15	57	**120**			172	
16	63		115			163
17	52	100		140		
18	48		88			
19	40					

Analysis Table

Columns	Size of item containing maximum frequency						
(1)			9				16
(2)			9	10		15	16
(3)		8	9				
(4)		8	9	10			
(5)			9	10	11		
(6)	7	8	9				
No. of times size occurs	1	3	6	3	1	1	2

The size 9 is, therefore, the mode.

PROBLEM 1.48. *The following table gives the length of life of 150 electri lamps :—*

Life (hours)	*Frequency of lamps*
0 to 400	*4*
400 to 800	*12*
800 to 1200	*40*
1200 to 1600	*41*
1600 to 2000	*27*
2000 to 2400	*13*
2400 to 2800	*9*
2800 to 3200	*4*

Calculate the arithmetic average, the median and the mode.

SOLUTION. Determination of Mode by grouping.

Life (hours)	Frequency of lamps					
	(1)	(2)	(3)	(4)	(5)	(6)
0- 400	4	}16		}56		
400- 800	12		}52		}93	
800-1200	40	}**81**				}**108**
1200-1600	**41**		}**68**	}**81**		
1600-2000	27	}40			}49	
2000-2400	13		}22			}26
2400.2800	9	}13				
2800-3200	4					

Columns	Size of group containing maximum frequency				
(1)			1200-1600		
(2)		800-1200	1200-1600		
(3)			1200-1600	1600-2000	
(4)			1200-1500	1600-2000	2000-2400
(5)	400-800	800-1200	1200-1600		
(6)		800-1200	1200-1600	1600-2000	
No. of times the size occurs	1	3	6	3	1

Therefore, 1200-1600 group is the modal group. Mode lies in this group and by applying the formula of interpolation, *viz.*,

$$Z = l_1 + \frac{f_1 - f_0}{f_1 - f_0 - f_2}(l_2 - l_1);$$

Where Z stands for Mode.

l_1 and l_2 stand for the lower and upper limits of the modal group,

f_1 stands for frequencies in the modal group.

f_0 ,, ,, ,, group preceding the modal group.

f_2 ,, ,, ,, group succeeding the modal group

We get, $Z = 1200 + \frac{41 - 40}{82 - 40 - 27} \times 400$

$= 1226.67.$ hours.

Thus,

The Modal life of the lamp = 1226.67 hours.

The Median life of the lamp = 1385.36 hours (size of 75th item).

And the Mean life of the lamp = 1453.33 hours.

(Mean and median may be calculated in the manner as explained before).

PROBLEM 1.49. *The following is the age distribution of candidates appearing at the Matriculation and Intermediate Arts examination of a University in 1974.*

Age in years	*12-*	*13-*	*14-*	*15-*	*16-*	*17-*	*18-*	*19-*	*20-*	*21-*	*22*	*Total*
Matriculation	*5*	*48*	*189*	*303*	*522*	*980*	*981*	*794*	*515*	*474*	×	*4811*
Intermediate	×	×	×	*5*	*45*	*87*	*127*	*150*	*155*	*127*	*175*	*871*

Compare the median and modal ages of the Matriculation candidates with those of I. A. candidates.

SOLUTION. Matriculation Examination.

Determination of Modal age by grouping

Age group	No. of Candidates (f)						No. of times a group contains maximum frequency	Cumulative Frequency
(m)	(1)	(2)	(3)	(4)	(5)	(6)		
12-13	5	53		242				5
13-14	48		237		540			53
14-15	189	492				1014		242
15-16	303		825	1805				545
16-17	522	1502			2483		1	1067
17-18	980		1961			2755	3	2047
18-19	981	1775		2290			6	3028
19-20	794		1309		1783		3	3822
20-21	515	989					1	4337
21-22	474							4811
22-23								4811

The modal age group is, therefore, 18-19 age group. By interpolation.

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_3 - l_1)$$

$$= 18 + \frac{981 - 980}{1962 - 980 - 794}\ (19 - 18) = 18 + \frac{1}{188}$$

= 18.005 years approx.

Median = age of $\frac{n}{2}$ *i. e.*, 2405.5th candidate

which lies in 18-19 age group ;

By interpolation,

$$\text{Median} = 18 + \frac{19 - 18}{981}\ (2405.5 - 2047) = 18.37 \text{ years approx.}$$

Intermediate Arts Examination.

Determination of modal age by grouping

Age Group	No. of candidates						No of times a group contains maximum frequency	Cumulative Frequency
	(1)	(2)	(3)	(4)	(5)	(6)		
12-13	×							×
13-14	×							×
14-15	×							×
15-16	5	50		137				5
16-17	45		132		259			50
17-18	87	214				364		137
18-19	127		277	432			1	264
19-20	150	305			432		3	414
20-21	155		282			457	5	569
21-22	127	302					3	696
22-23	175						2	871

The modal age group is therefore 20-21 age group,

By interpolation,

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

$$= 20 + \frac{155 - 150}{310 - 150 - 127}(21 - 20) = 20.15 \text{ years}$$

Median age = the age of the 435.5th candidate,

which lies in 20-21 age group.

By interpolation,

$$= 20 + \frac{21 - 20}{155}(435.5 - 414) = 20.14 \text{ years.}$$

On comparison we find that the median and modal ages of Matriculation candidates are lower than those of I. A. candidates by about 2 years.

PROBLEM 1.50. *The following table shows the frequency with which profits are made. What is the Mode ?*

	Frequency
Exceeding Rs, 3,000 and not exceeding 4,000	*83*
Exceeding Rs. 4,000 and not exceeding 5,000	*27*
Exceeding Rs. 5,000 and not exceeding 6,000	*25*
Exceeding Rs. 6,000 and not exceeding 7,000	*50*
Exceeding Rs. 7,000 and not exceeding 8,000	*75*
Exceeding Rs. 8,000 and not exceeding 9,000	*38*
Exceeding Rs. 9,000 and not exceeding 10,000	*18*

SOLUTION. Determination of mode by grouping.

Profits in rupees	Frequency						No. of items group occurs in the maximum
	(1)	(2)	(3)	(4)	(5)	(6)	
Exceeding- Not-exceeding							
3000-4000	83	110		135			1
4000-5000	27		52		102		
5000-6000	25	75				150	1
6000-7000	50		125	163			3
7000-8000	75	113			131		5
8000-9000	38		56				3
9000-10000	18						1

The Mode therefore lies in the profit group exceeding Rs. 7,000 and not exceeding Rs. 8,000

By interpolation, The $\text{Mode} = l_1 + \dfrac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$

$$= 7000 + \frac{75-50}{150-50-38}(8000-7000)$$

$$= 7403 \text{ rupees approx.}$$

Thus, the modal profit as revealed by this series is Rs. 7403 approx.

PROBLEM 1.51. *Find the modal wage group from the following table :—*

Wages in Rupees	*No. of labourers*
Above 330	*520*
,, *340*	*470*
,, *350*	*399*
,, *360*	*210*
,, *370*	*105*
,, *380*	*45*
,, *390*	*7*

SOLUTION. Determination of modal wage group by grouping.

Wage-groups Rupees	No. of labourers (frequency)						No. of times a group contains maximum frequency
	(1)	(2)	(3)	(4)	(5)	(6)	
Exceeding Not-exceeding							
330-340	50	121		310			1
340-350	71		260		365		3
350-360	**189**	294				354	6
360-370	105		165	203			
370-380	60	98			105		3
380-390	38		45				1
390 and above	7						

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_1}\,(l_2 - l_1) = 350 + \frac{189 - 71}{378 - 71 - 105}\,(360\text{-}350)$$

$$= 350 + 5.84 = 355.84 \text{ Rupees,}$$

The modal wage group is therefore, the wage group which is exceeding Rs. 350 but not Rs. 360.

PROBLEM 1.52. *Find out the median and the mode for the following table :—*

No. of days absent		*No. of students*
Less than	*5*	*29*
Less than	*10*	*224*
Less than	*15*	*465*
Less than	*20*	*582*
Less than	*25*	*634*
Less than	*30*	*644*
Less than	*35*	*650*
Less than	*40*	*653*
Less than	*45*	*655*

SOLUTION. Determination of mode by grouping.

No. of days absent	No. of students (Frequency)						No. of times group occurs in the maximum.	Cumulative Frequency
	(1)	(2)	(3)	(4)	(5)	(6)		
0- 5	29	244		465			1	29
5-10	195		436		553		3	224
10-15	241	358				410	6	465
15-20	117		169	179			3	582
20-25	52	62			68		1	634
25-30	10		16			19		644
30-35	6	9		11				650
35-40	3		5					653
40-45	2							655

The modal group is, therefore, the 10-15 group. By interpolation,

$$\text{The Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}\,(l_2 - l_1)$$

$$= 10 + \frac{241 - 195}{482 - 195 - 117}\,(15 - 10) = 11.35 \text{ days.}$$

Median = the no. of days the 327.5th student remained absent which lies in 10-15 days group.

$$\text{By interpolation} = 10 + \frac{15 - 10}{241}\,(327.5 - 224) = 12.15 \text{ days.}$$

PROBLEM 1.53. *Find the modal wage from the following data :*

Weekly		*Wage*	*No. of Wage earners*
£ P		*£ P*	
12 60	*to*	*17 60*	*4*
17 60	*to*	*22 60*	*44*
22 60	*to*	*27 60*	*38*
27 60	*to*	*32 60*	*28*
32 60	*to*	*37 60*	*6*
37 60	*to*	*42 60*	*8*
42 60	*to*	*47 60*	*12*
47 60	*to*	*52 60*	*2*
52 60	*to*	*57 60*	*2*

SOLUTION. Location of mode by grouping.

Weekly wage	No. of wage-earners (Frequency)						No. of times group occurs in the maximum
	(1)	(2)	(3)	(4)	(5)	(6)	
£ P £ P							
12 60 to 22 60	48	**114**		**128**			2
22 60 to 32 60	**66**		**80**		**94**		5
32 60 to 42 60	14	28				**30**	4
42 60 to 52 60	14		16				2
52 60 & over	**2**						1

The modal wage group is therefore the 22.60 – 32.60 £ wage group.

By interpolation,

$$\text{Modal wage} = 22.6 + \frac{66-48}{132-48-14}\,(32.6-22.6)$$

$$= 25.17 \text{ sterling.}$$

PROBLEM 1.54. *Find out the mode of the following series :—*

Size of item	*Frequency*
0- 9.99	*10*
10-19·99	*14*
20-29.99	*16*
30-39.99	*14*
40-49.99	*11*
50-59.99	*13*
60-69.99	*17*
70-79.99	*13*

SOLUTION Location of mode by grouping.

Size of item	Frequency (1)	(2)	(3)	(4)	(5)	(6)	Number of times a group occurs in the max. Frequency
0-9.99	10	24		40			1
10-19.99	14		30		44		3
20-29.99	16	30				41	4
30-39.99	14		25	38			2
40-49.99	11	24			41		0
50-59.99	13		30			43	2
60-69.99	**17**	30					4
70-79.99	13						2

On analysing the above table we find that there are two modal groups, *viz.* 20-29.99 and 60-69.99. This is a case of a bi-modal series. In such cases it is better to compute some other average than mode.

PROBLEM 1.55 *From the data given below find the mode.*

Ages	*20-25*	*25-30*	*30-35*	*35-40*	*40-45*	*45-50*	*50-55*	*55-60*
No. of persons	*50*	*70*	*80*	*180*	*150*	*120*	*70*	*50*

SOLUTION. Location of mode by grouping

Class interval (m)	Frequency (1)	(2)	(3) (f)	(4)	(5)	(6)
20-25	50	120		200		
25-30	70		150		330	
30-35	80	260				**410**
35-40	**180**		**330**	**450**		
40-45	150	**270**			**340**	
45-50	120		190			240
50-55	70	120		120	50	
55-60	50					

ANALYSIS TABLE

Columns		Size of group containing maximum frequency			
1		35-40			
2			40-45	45-50	
3		35-40	40-45		
4		35-40	40-45	45-50	
5			40-45	45-50	50-55
6	30-35	35-40	40-45		
No. of times the size occurs	1	4	5	3	1

Therefore, 40-45 group is the modal group.

By interpolation,

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_0} \times (l_2 - l_1)$$

$$= 40 + \frac{150 - 180}{2 \times 150 - 180 - 120} \times 5$$

$$= 40 + \frac{-30}{0} \times 5$$

Division by zero is not a permissible operation, Hence, in this case mode cannot be determined by this method.

So $\text{Mode} = l_1 + \frac{f_2}{f_0 + f_2} \times i$ wherein i represent $(l_2 - l_1)$

$$= 40 + \frac{120}{180 + 120} \times 5$$

$$= 40 + 2 = 42 \text{ years},$$

PROBLEM 1.56. *The following table gives the monthly income of 24 families in a certain locality :—*

Serial No. of the family	*Monthly income in rupees*	*Serial No. of the family*	*Monthly income in rupees*
1	*260*	*13*	*296*
2	*500*	*14*	*298*
3	*286*	*15*	*304*
4	*295*	*16*	*275*
5	*300*	*17*	*280*
6	*350*	*18*	*294*
7	*310*	*19*	*300*
8	*274*	*20*	*275*
9	*290*	*21*	*700*
10	*292*	*22*	*282*
11	*480*	*23*	*400*
12	*380*	*24*	*284*

Calculate the arithmetic average, the median and the mode of the above incomes. Which average would represent the above series the best ? Give reasons.

SOLUTION. The total of all the 24 items is Rs. 8005

$\therefore$ Arithmetic average $= \frac{8005}{24}$ rupees $= 333.54$ rupees.

If these items are arranged in ascending order of magnitude the values of the 12th and 13th items would be 295 and 296 respectively.

$\therefore$ Median $= \frac{295 + 296}{24}$ rupees or 295.5 rupees which is the value of the middle item.

For the calculation of mode the data would have to be grouped in class intervals as has been done below :—

Income Rs.	No. of families
260-280	4
280-300	10
300-320	4
320-340	0
340-360	1
360-380	0
380-400	1
400 and above	4

Mode lies in the 280—300 group. By interpolation

$$Z=l_1+\frac{f_1-f_0}{2f_1-f_0-f_2}(l_2-l_1)$$

$$=280+\frac{10-4}{20-4-4}(300-280)=290 \text{ rupees.}$$

Since 18 families out of 24 earn less than Rs. 320, arithmetic average doe not best represent the series. Median represents the series best because the value of mode has been interpolated and such values are not very accurate.

PROBLEM 1.57. *Number of workers and average earnings per worker i four factories in 1969 and 1971 were as under :—*

Year	*Factory*	*No. of workers*	*Average Earnings in Rupees*
1969	*A*	*300*	*143*
	B	*200*	*139*
	C	*50*	*110*
	D	*150*	*120*
1971	*A*	*200*	*153*
	B	*500*	*141*
	C	*100*	*113*
	D	*200*	*125*

Calculate the average earnings per worker in 1969 and 1971 for all the fou factories combined.

SOLUTION. Calculation of combined mean

For 1971 :

$$\text{Combined mean}=\frac{n_1m_1+n_2m_2+n_3m_3+n_4m_4}{n_1+n_2+n_3+n_4}$$

For 1969

$$\text{Combined mean or } m_{12}=\frac{(300\times143)+(200\times139)+(50\times110)+(150\times12}{300+200+50+150}$$

$$\text{or } m_{12}=\frac{42{,}900+27{,}800+5{,}500+18{,}000}{700}$$

$$\text{or } m_{12}=\frac{94{,}200}{700}=134.57 \text{ rupees approx.}$$

For 1971

$$\text{Combined mean or } m_{12}=\frac{(200\times153)+(500\times141)+(100\times113)+(200\times1}{200+500+100+200}$$

$$=\frac{30{,}600+70{,}500+11{,}300+25{,}000}{1{,}000}$$

$$=\frac{1{,}37{,}400}{1{,}000}=137.4 \text{ rupees.}$$

PROBLEM 1.58. *The mean wage of 100 labourers working in a factory r ning two shifts of 70 and 80 workers respectively is Rs. 284. The mean wage of 70 labourers working in morning shifts is Rs. 290. Find the mean wage of workers working in the evening shift.*

SOLUTION.

We know that combined mean$=\dfrac{n_1m_1+n_2m_2}{n_1+m_2}$

$m_{12}=$Rs. 284, $n_1=70$, $n_2=30$, $m_1=290$

$$284=\frac{70\times290+30\times m_2}{100}$$

$$28400=20300+30m_2$$

$$30m_2=8100$$

$$\therefore\ m_2=\frac{8100}{30}=270 \text{ rupees.}$$

Thus the mean wage of 30 workers working in the evening shift is Rs. 27

PROBLEM 1.59. *The mean salary paid to 1000 employees of a factory was found to be Rs. 180.4. Later on it was discovered that the wages of two employees were wrongly taken as 297 and 165 instead of 197 and 185. Find the correct mean.*

SOLUTION. Calculation of correct mean

$$\text{Mean}=\frac{\Sigma m}{n}$$

$$180.4=\frac{\Sigma m}{1000}$$

$$\therefore\ \Sigma m=1,80,400$$

$$\text{Correct } \Sigma m=1,80,400-297-165+197+185=1,80,320$$

$$\text{Correct mean}=\frac{1,80,320}{1,000}=\text{Rs. } 180.32$$

PROBLEM 1.60. *The average marks for arithmetic in a class of 30 was 52. The top six students had an average of 31. What was the average mark of the other students ?*

SOLUTION. Combined arithmetic mean$=\dfrac{n_1m_1+n_2m_2}{n_1+n_2}$

Where n_1 and n_2 stand for the number of items in the two series and m_1 and m_2 for the mean of the two series respectively.

By substituting the values given in the problem, we get

$$52=\frac{6\times31+24\times m_2}{6+24}$$

$$\text{or}\quad 52=\frac{186+24m_2}{30}$$

$$\text{or}\quad 1560=186+24m_2$$

$$\therefore\ 24\,m_2=1374$$

$$\therefore\quad m_2=\frac{1374}{24}=57.25 \text{ marks.}$$

Hence, the average marks of the other students is 57.25.

PROBLEM 1.61. *The mean age of a combined group of men and women is 30 years. If the mean age of the group of men is 32 and that of the group of women is 27, find out the percentage of men and women in the group.*

SOLUTION. Let men be $x\%$; then females would be $(100\text{-}x)\%$

Total age in years of the whole group (let it be 100) and that of men and women are as follows—

$$32x+27\,(100\text{-}x)=30\times100$$

$$\text{or}\quad 32x+2,700-27x=3,000$$

$$\text{or}\quad 32x-27x=3,000-2,700$$

$$\text{or}\quad 5x=300$$

$$\therefore\ x=\frac{300}{5}=60$$

It means there were 60% men and 100—60=40% women.

PROBLEM 1.62. *In a moderately skewed distribution*

(*a*) *Arithmetic mean=24.6 and the mode =26.1. Find the value of the median and explain the reason for the method employed.*

(*b*) *In a moderately asymmetrical distribution the value of median is 42.8 and the value of mode is 40. Find the mean.*

(*c*) *If the mode and mean of a moderately asymmetrical series are respectively 26 and 20.2 meters compute most probable median.*

SOLUTION. All the three cases can be solved with the help of the following formula,

$$(\text{Median}-\text{Mode})=\frac{2}{3}(\text{Mean}-\text{Mode})$$

$$\text{Mode}=3\ \text{Median}-2\ \text{Mean}$$

(*a*) Substituting the given values, we get

$$26.1=3\ \text{Median}-2\times 24.6$$

or $26.1=3\ \text{median}-49.2$

$$\therefore\ \text{Median}=\frac{75.3}{3}=25.1$$

(*b*) Substituting the given values, we get

$$40=3\times 42.8-2\ \text{Mean}$$

or $40=128.4-2\ \text{Mean}$

$$\therefore\ \text{Mean}=\frac{88.4}{2}=44.2$$

(*c*) Substituting the given values, we get

$$26=3\ \text{Median}-2\times 20.2$$

or $26=3\ \text{Median}-40.4$

$$\therefore\ \text{Median}=\frac{66.4}{4}=22.13\ \text{meters approx.}$$

PROBLEM 1.63. *How are the mean and median affected when it is known that for a group of 10 students, scoring an average of 60 marks, the best pape was wrongly marked 80 instead of 75 ?*

SOLUTION. Calculation of correct mean

$$\text{Arithmetic mean or} = a\ \frac{\Sigma \text{m}}{n}$$

$$\text{or } 60=\frac{\Sigma \text{m}}{10}$$

$$\text{or } \Sigma \text{m}=600$$

Correct $\Sigma \text{m}=600-80+75=595$

$$\text{Correct mean}=\frac{595}{10}=59.5\ \text{marks}$$

The value of median will not be affected by this error because this is positional average and the median value is 60 which is far away from the values 80 and 75.

PROBLEM 1.64. *An incomplete distribution is given as follows:*

Variable	*Frequency*
10-20	*12*
20-30	*30*
30-40	*?*
40-50	*65*
50-60	*?*
60-70	*25*
70-80	*18*
	229

You are given that the median value is 46.

(*a*) *Using the median formula fill up missing frequencies.*

(*b*) *Calculate the arithmetic mean of the computed table.*

SOLUTION Calculation of missing frequencies and mean.

Variable (*m*)	Frequency (*f*)	Cumulative frequency (*cf*)	Midvalues (*m.v.*)	Deviations from as av. (45)	Stepdeviation (*dx*)	Total deviations (*fdx*)	Frequency (*f*)
10-20	12	12	15	—30	—3	—36	12
20-30	30	42	25	—20	—2	—60	30
30-40	x	$42+x$	35	—10	—1	—34	34
40-50	65	$107+x$	45	0	0	0	65
50-60	$(79\text{-}x)$	186	55	+10	+1	+45	45
60-70	25	211	65	+20	+2	+50	25
70-80	18	229	75	+30	+3	+54	18
	$n=229$					$\Sigma fdx=+19$	$n=229$

Let the frequency of the group 30-40 be assumed x

frequency of the group 50-60$=229-(150+x)=79—x$

For median number we shall use $\frac{1}{2}(n+1)$ and not $n/2$ because the given median has been calculated as such.

$m=l_1+\dfrac{l_2-l_1}{f}(m—c)$, by putting the known values

$$46=40+\frac{50-40}{65}\{115-(42+x)\}$$

$$\text{or}\quad 46=40+\frac{10}{65}(115-42-x)$$

$$\text{or}\quad 6=\frac{10}{65}\cdot(73-x)$$

$$\text{or}\quad 390=730-10x$$

$$\therefore\ x=\frac{340}{10}=34$$

Hence, first missing frequency is 34 and second missing frequency is 45.

Arithmetic average or $a=45+\dfrac{19\times10}{229}=45.83$ approx.

PROBLEM 1.65. *The median and mode of the following wage distribution are known to be Rs. 233.5 and Rs. 234 respectively. Three frequencies Values from the table are, however, missing. Can you find these frequency Values?*

Wages in Rs.	*Frequencies*
200-210	*4*
210-220	*16*
220-230	*?*
230-240	*?*
240-250	*?*
250-260	*6*
260-270	*4*
	230

SOLUTION . Let the frequencies in the following wage groups be as noted against them.

220-230	x
230-240	y
240-450	$230—30—x—y$

The median is 233.5 (given) hence it lies in the 230-240 class interval. Mode is 234 (given). It also lies in the same class interval.

$$\text{Median} = l_1 + \frac{l_2 - l_1}{f}(m\text{-}c)$$

$$233.5 = 230 + \frac{240-230}{y}\left\{\frac{230}{2} - (20+x)\right\}$$

or $$3.5 = \frac{10}{y}(115-20-x)$$

or $$3.5y = 1150-200-10x$$

or $$10x+35y=950 \qquad (i)$$

$$\text{Mode} = l_1 + \frac{f_1-f_0}{2f_0-f_1-f_2}\times(l_2-l_1)$$

or $$234 = 230 + \frac{y-x}{2y-x-(200-x-y)}\times 10$$

or $$4 = \frac{y-x}{2y-x-200+x+y}\times 10$$

or $$4 = \frac{10y-10x}{3y-200}$$

or $$12y-800 = 10y-10x$$

or $$10x+2y=800 \qquad (ii)$$

By subtracting (*ii*) from (*i*), we get

$$1.5y=150$$

$$\therefore y=100$$

Substituting the Value of y in (*i*) we get

$$10x+350=950$$

or $$10x=600$$

$$\therefore \quad x=60$$

Hence the missing frequencies are

Class intervals	Frequencies
220-230	60
230-240	100
240-250	40

GEOMETRIC MEAN

PROBLEM 1.66. *Calculate the simple geometric mean from the following items :—*

133, 141, 125, 173, 182,

SOLUTION. Calculation of the geometric mean.

Size of item	Logarithms
133	2.1239
141	2.1492
125	2.0969
173	2.2380
182	2.2601
$n=5$	$\Sigma\text{logs}=10.8681$

According to the formula, *viz.*

$$\text{Geometric mean} = \sqrt[n]{a\times b\times c\times d\ldots\ldots\ldots\times n}$$

$$= \text{Anti log}\left[\frac{\log a+\log b+\log c+\log n}{n}\right]$$

$$= \text{Anti log}\left[\frac{10.8681}{5}\right]$$

$$= \text{Anti log } 2.1736$$

$= 149$ (to the nearest whole number)

Thus the geometric mean is **149.**

Alternate Method :

Size of item	Logs	Deviations from assumed log mean (2.000) (dx)
133	2.1239	.1239
141	2.1492	.1492
125	2.0969	.0969
173	2.2380	.2380
182	2.2601	.2601
$n=5$		$\Sigma dx=.8681$

$$\text{Geometric Mean}=\text{Anti log}\left[\text{assumed log}+\frac{\Sigma\text{ Deviation}}{n}\right]$$

$$=\text{Anti log}\left[2+\frac{\cdot 8681}{5}\right]$$

$$=\text{Anti log } 2.1736$$

$$=149 \text{ (to the nearest whole number)}$$

Thus the geometric mean is 149.

PROBLEM 1.67. *Calculate the geometric mean of the following figures :—*

5. 10, 192, 14374, 20498, 120674, 15491

SOLUTION. Calculation of the geometric mean.

Size of items	Logarithms
5	0.6990
10	1.0000
192	2.2833
14374	4.1575
20498	4.3118
120674	5.0816
15491	4.1903
$n=7$	$\Sigma\text{logs}=21.7235$

According to the formula, *viz.*

$$\text{Geometric mean}=\text{Anti log}\left(\frac{\log a+\log b+\ldots+\log n}{n}\right)$$

$$=\text{Anti log}\frac{\Sigma\text{logs}}{n}$$

$$\text{we get, } g=\text{Anti log}\left(\frac{21.7235}{7}\right)$$

$$=\text{Anti log } 3.1034$$

$$=1269.0$$

Thus the geometric mean is 1269.

Alternative Method

Size of items	Logs	Deviation from assumed log mean (3.0000). (dx)
5	0.6990	3.6990
10	1.0000	2.0000
192	2.2833	1.2833
14374	4·1574	1.1575
20498	4.3118	1.3118
120674	5.0816	2.0816
15491	4.1903	1.1903
$n=7$		$\Sigma dx=0.7235$

$$\text{Geometric Mean} = \text{Anti log}\left(3 + \frac{0.7235}{7}\right)$$
$$= \text{Anti log. } (3.1034)$$
$$= 1269.0$$

PROBLEM 1.68. *Calculate the geometric mean of the following two series :—*

(a)	(b)
2574	.8974
475	.0570
75	.0081
5	.5677
.8	.0001
.08	.0284
.005	.0854
.0009	.5672

SOLUTION Calculation of geometric mean

	Series A		Series B	
	Size of items	Logarithms	Size of items	Logarithms
(*a*)	2574	3·4106	.8974	$\bar{1}.9530$
(*b*)	475	2.6767	.0570	$\bar{2}.7559$
(*c*)	75	1.8751	.0081	$\bar{3}.9085$
(*d*)	5	0.6990	.5677	$\bar{1}.7542$
(*e*)	.8	$\bar{1}.9031$	.0002	$\bar{4}.3010$
(*f*)	.08	$\bar{2}.9031$	.0984	$\bar{2}.9930$
(*g*)	.005	$\bar{3}.6990$	.0854	$\bar{2}.9315$
(*h*)	.0009	$\bar{4}.9542$	.5672	$\bar{1}.7538$
		Σlogs=2.1208		Σlogs=10.3509

$$\text{Geometric Mean} = \text{Anti log}\left[\frac{\log a + \log b + \log c + \ldots + \log n}{n}\right]$$
$$= \text{Anti log}\left[\frac{\Sigma \text{logs}}{n}\right]$$

Series A

According to the formula, we have

$$= \text{Anti log}\frac{2.1208}{8} = \text{Anti log } .2651 = 1.841.$$

Series B

$$g = \text{Anti log } \frac{16 + 6.3509}{8} = \text{Anti log } 2.7939 = {\cdot}06222$$

PROBLEM 1.69. *Compute the weighted geometric average of Relative Prices of the following commodities for the year 1973 (Base year 1972- Price 100):—*

Commodity	*Relative Price*	*Weight (value produced in 1972)*
Corn	*128.8*	*1385*
Cotton	*62.4*	*819*
Hay	*117.7*	*842*
Wheat	*99.0*	*561*
Oats	*130.9*	*408*
Potatoes	*143.5*	*194*
Sugar	*125.6*	*142*
Barley	*150.2*	*100*
Tobacco	*101.1*	*103*
Rye	*116.2*	*25*
Rice	*117.5*	*17*
Oilseeds	*78.7*	*29*

How does it differ from the unweighted geometric mean, and why ?

SOLUTION. Computation of weighted geometric average of relative price.

Commodity	Relative Price	Log. of price	Weight	Weighted logs $w \times \log$.
Corn	128.8	2.1100	1385	2919.6500
Cotton	62.4	1.7952	819	1470.2688
Hay	117.7	2.0708	842	1743.6136
Wheat	99.0	1.9956	561	1119.5316
Oats	130.9	2.1168	408	863.6544
Potatoes	143.5	2.1568	194	418.4192
Sugar	125.6	2.0990	142	298.0580
Barley	150.2	2.1767	100	217.6700
Tobacco	101.1	2.0047	103	206.4841
Rye	116.2	2.0653	25	51.6325
Rice	117.5	2.0701	17	35.1917
Oilseeds	78.7	1.8960	29	54.9840
		Σlogs 24.5570	$W=4625$	$\Sigma(\log \times W)=9399.1579$

Simple geometric average of relative prices. $=\text{Anti log}\left[\dfrac{\Sigma \text{logs}}{n}\right]$

$$=\text{Anti log}\,\frac{24.5570}{12}$$

$$=\text{Anti log } 2.0464=111.3$$

Weighted geometric mean of relative prices.

$$=\text{Anti log}\left(\frac{(\log a \times w_1+\log b \times w_2+\ldots+\log n \times wn)}{w_1+w_2+\ldots wn}\right)$$

$$=\text{Anti log}\frac{\Sigma(\log \times w)}{\Sigma(w)}=\text{Anti log}\ \frac{9399{\cdot}1579}{4625}$$

$$=\text{Anti log } 2.0323=107.7$$

PROBLEM 1.70. *Compute the geometric mean of the following series.*

Marks	*No. of Students*
0-10	*5*
10-20	*7*
20-30	*15*
30-40	*25*
40-50	*8*
	60

SOLUTION. Computation of the geometric mean of the marks of 60 students.

Marks	Mid-points	Log	Frequency (f)	Product of Log and f
0-10	5	0.6990	5	3.4950
10-20	15	1.1761	7	8.2327
20-30	25	1.3979	15	20.9685
30-40	35	1.5441	25	38.6025
40-50	45	1.6532	8	13.2256
			60	84.5243

$$\text{G. M.}=\text{Anti log}\ \frac{84.5243}{60}=\text{Anti log } 1.40874=25.64 \text{ marks.}$$

PROBLEM 1.71. *The price of a commodity increased by 5 % from 19 to 1969, 8 % from 1969 to 1970 and 77% from 1970 to 1971. The avera increase from 1968 to 1971 is quoted as 26 % and not 30 % Explain and veri the arithmetic average.*

SOLUTION. The appropriate average here is the geometric mean and n the arithmetic mean. The arithmetic mean of 5, 8 and 77 is 30 but this is n the correct answer. Correct answer shall be obtained if we calculate geometr mean.

% rise	Price at the end of the year taking preceding year as 100 (m)	log (m)
5	105	2.0212
8	108	2.0334
77	177	2.2480
	$n=3$	$\Sigma\log(m)=6.302$

$$\text{G. M}=\text{Anti log}\,\frac{\Sigma\log m}{n}$$

$$=\text{A. L.}\,\frac{6.3026}{3}$$

$$=\text{A. L. } 2.100\,9=126.2$$

The average increase from 1968 to 1971 = 126.2 − 100 = 26.2% or 26 approx.

VERIFICATION. *When the average rise is 30%*

Year	Rate of change	Total change	Price at the end of each year
1 year	30% on 100	30	130.0
2 year	30% on 130	39	169.0
3 year	30% on 169	50.7	219.7

When the average rise is 26%

Year	Rate of change	Total change	Price at the end of each year
1 year	26% on 100	.26	126
2 year	26% on 126	32.76	158.76
3 year	26% on 158.76	41.26	200.02

When the rise of 5, 8 and 77%, the changed price at the end of each year :

1 year	5% on 100	5.00	105.0
2 year	8% on 105	8.40	113.4
3 year	77% on 113.4	87.32	200.7

The above calculations make it clear that in the second and third cases th price at the end of the third year is almost the same, the slight difference bein due to approximation. Hence, the average rise in price is 26%.

PROBLEM 1.72. *The population of a country was 300 million in 1951. became 520 million in 1969. Calculate the percentage compound rate of grow per annum.*

SOLUTION Let r be the rate of growth. Applying the compoun interest formula : $P_n=P_0\,(1+r)^n$

$$\text{or } 520=300\,(1+r)^{18}$$

Taking logarithms

$$\log\ (1+r)=\frac{\log 520-\log 300}{18}$$

$$=\frac{2.7160-2.4771}{18}=0.0133$$

$$1+r=1.0310$$

$$r=.0310=3.1\%$$

PROBLEM 1.73. *The population of a country increased by 20 % in the first decade, 30 % in the second decade and 45 % in the third decade. What is the average rate of increase per decade in the population ?*

SOLUTION. Calculation of average rate per decade

Years	% rise	Population at the end of the decade m	Log of population
1st decade	20	120	2.0792
2nd decade	30	130	2.1139
3rd decade	45	145	2.1614
			Σlog=6.3545

$$\text{G.M.}=\text{A. L.}\left(\frac{\Sigma\log m}{n}\right)$$

$$=\text{A. L.}\left(\frac{6.3545}{3}\right)$$

$$=\text{A. L. } 2.1182$$

$$=131.3$$

∴Average increase per decade is 131.3 − 100 = 31.3 %

PROBLEM 1.74. *The number of divorces per 1000 marriages in the United States increased from 84 in 1950 to 108* in *1970. Find the annual rate of increase of the divorce rate for the period 1950 to 1970.*

SOLUTION. Let r be the rate of growth.

Applying the compound interest formula :

$$Pn=Po(1+r)n$$

$$108=84\ (1+r)\ 20$$

Taking logarithms

$$\log\ (1+r)=\frac{\log 108-\log 84}{20}$$

$$\log\ (1+r)=\frac{2.0334-1.9243}{20}$$

$$=\frac{0.1091}{20}=0.0055$$

$$1+r=1.013$$

$$r=0.0130=1.3\%$$

PROBLEM 1.75. *The annual rates of growth of a factory in 5 years are 5. 0; 7. 5; 2. 5; 5. 0 and 10 per cent respectively. What is the compound rate of growth of output per annum for the period ?*

SOLUTION. Applying the geometric mean

Annual rate of growth	Output Relative at the end of the year m	$\log m$
5.0	105.0	2.0212
7.5	107.5	2.0314
2.5	102.5	2.0107
5.0	105.0	2.0212
10.0	110.0	2.0414
$n=5$		$\Sigma \log m=10.1259$

$$\text{Geometric Mean}=\text{Antilog}\ \frac{\Sigma \log m}{n}$$

$$=\text{Anti log}\frac{10.1259}{5}$$

$$=\text{Antilog } 2.0252$$

$$=105.9$$

The compound rate of growth of output per annum for the period is 105.9 —100=5.9%.

HARMONIC MEAN

PROBLEM 1.76. *The monthly income of fifteen families is given below in Rupees :—*

80, 2,500, 90, 1,200, 1,450, 7,200, 120, 1,060, 150. 480, 360, 96, 200, 520, 60.

Calculate the Harmonic Mean

SOLUTION. Computation of the Harmonic Mean of the monthly incomes of fifteen families

Size of items Rs.	Reciprocals
80	.01250
2500	.00040
90	.01111
1200	.00083
1450	.00069
7200	.00014
120	.00833
1060	.00094
150	.00667
480	.00208
360	.00278
96	.01042
200	.00500
520	.00192
60	.01667
	.08048

$$\text{Harmonic Mean} = \text{Reciprocal}\ \frac{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\ldots\ldots+\frac{1}{n}}{n}$$

where $a, b, c, \ldots\ldots n$ represent the values of the n items of the variable, and n is the number of items.

$$=\text{Reciprocal}\frac{\Sigma \text{ Reciprocals}}{n}$$

Substituting the values,

$$\text{Harmonic Mean of the annual incomes} = \text{Reciprocal } \frac{.08048}{15}$$

$$= \text{Reciprocal } .00536$$

$$= 186.5 \text{ rupees.}$$

PROBLEM 1.77. *Calculate the Harmonic Mean of the following items.*
1.0; 1.5; 5.0; 15.0; 250.0; .5; .05; .095; 1245.0; .009;

SOLUTION. Calculation of the Harmonic Mean

Size of the items	Reciprocals
1.0	1.0000
1.5	.6667
5.0	.2000
15.0	.0666
250.0	.0040
.5	2.0000
.05	20.0000
.095	10.5300
1245.0	.0008
.009	111.0000
	145.4681

$$\text{Harmonic mean} = \text{Reciprocal} \frac{\Sigma \text{ (Reciprocal) of the items)}}{n}$$

$$= \text{Reciprocal } \frac{145.4681}{10}$$

$$= \text{Reciprocal } 14.54681 = .06897.$$

PROBLEM 1.78. *Calculate the weighted harmonic average of the following items:*

Items	*Weight*
1	*5*
.5	*10*
10.0	*20*
45.0	*10*
175.0	*15*
.01	*5*
4.0	*15*
11.2	*8*

SOLUTION. Computation of the Harmonic Mean.

Item	Reciprocal	Weight	Weight × Reciprocals
1.0	1.0000	5	5.0000
.5	2.0000	10	20.0000
10:0	.1000	20	2.0000
45.0	.0222	10	.2220
175.0	.0057	15	.0855
.01	100.0000	2	200.0000
4.0	.2500	15	3.7500
11.2	.0893	8	.7144
		85	231.7719

$$\text{The Weighted Harmonic Mean} = \text{Reciprocal } \frac{231.7719}{83}$$

$$= \text{Reciprocals } 2.727 = .3663.$$

PROBLEM 1.79. *The following table gives the marks (out of 10) obtained by a group of students. Calculate the Harmonic Mean of the series.*

Marks	*Number of Students*
1	*7*
2	*11*
3	*17*
4	*17*
5	*26*
6	*31*
7	*11*
8	*1*
9	*1*
	121

SOLUTION. Calculation of the Harmonic Mean.

Marks (m)	Number of Students (f)	f/m
1	7	7.000
2	11	5.400
3	17	5.667
4	17	4.250
5	26	5.200
6	31	5.167
7	11	1.571
8	1	0.125
9	1	0.111
	$\Sigma f=121$	$\Sigma f/m=34.591$

$$\text{Harmonic mean}=\text{Reciprocal}\left[\frac{\Sigma f/m}{\Sigma f}\right]$$

$$=\text{Reciprocal}\,\frac{34.591}{121}$$

$$=\text{Reciprocal } 0.2858=3.519 \text{ marks.}$$

Note. f/m is equal to reciprocal of 'm' multiplied by 'f' or reciprocal of an item multiplied by its weight.

PROBLEM 1.80. *A man travelled by auto for 3 days. He covered 480 kilometers each day. He drove the first day 10 hours at 48 kilometers per hour, the 2nd day 12 hours at 40 kilometers per hour and the third day 15 hours at 32 kilometers per hour. What was his average speed ?*

SOLUTION. The problem will be solved with the help of Harmonic mean.

$$\text{H.M.}=\frac{N}{\Sigma\left(\frac{1}{m}\right)}$$

$$=\frac{3}{\frac{1}{48}+\frac{1}{40}+\frac{1}{32}}$$

$$=\frac{3}{37/480}=\frac{3}{1}\times\frac{480}{37}=38.92 \text{ kilometres per hour}$$

PROBLEM 1.81. *A man travels from one city to another. The distance between the two cities is 4 kilometres. He drives his car at 40 kilometre per hour. After travelling one kilometre the car stops running. He then travels on a tonga at 10 kilometres per hour. After travelling a distance of 1.5 kilometres, he leaves the tonga and covers the remaining distance on foot at 4 kilometres per hour. Find the average speed per hour of that person.*

SOLUTION. Calculation of weighted harmonic mean.

$$\text{H.M.}\omega = \frac{4}{\left(\frac{1}{40}\times 1\right)+\left(\frac{1}{10}\times 1.5\right)+\left(\frac{1}{4}\times 1.5\right)}$$

$$= \frac{4}{\frac{1}{4}+\frac{1.5}{10}+\frac{1,5}{4}}$$

$$= \frac{4}{22/40} = \frac{4}{1}\times\frac{40}{22} = 7.27 \text{ kilometres per hour approx.}$$

PROBLEM 1.82. *A man travels 20 kilometres at 40 kilometres per hour, 10 kilometres at 30 kilometres per hour and 30 kilometres at 60 kilometres per hour. What was his average speed ?*

SOLUTION. Weighted harmonic mean would be appropriate here.

$$\text{Harmonic Mean}\omega = \frac{\Sigma w}{\left(\frac{1}{a}\times w_1\right)+\left(\frac{1}{b}\times w_2\right)+\left(\frac{1}{c}\times w_3\right)}$$

The various speeds are 40 km.p.h., 30 km.p.h., and 60 km.p.h., hence a, b and c respectively are 40, 30 & 60. The distances covered are respectively 20, 10 and 30 kilometres

Hence w_1, w_2 and w_3 are 20, 10 and 30.

$$\text{H.M.}\omega = \frac{60}{\left(\frac{1}{40}\times 20\right)+\left(\frac{1}{30}\times 10\right)+\left(\frac{1}{60}\times 30\right)}$$

$$= \frac{60}{\frac{1}{2}+\frac{1}{3}+\frac{1}{2}} = \frac{60}{1}\times\frac{6}{8} = 45 \text{ Km. p.h.}$$

Thus the average speed is 45 km. p.h.

Verification

Speed (km. p.h.)	Distance (kilometres)	Time taken (in minutes)
40	20	30
30	10	20
60	30	30
	60	80

In 80 minutes the man covers 60 kilometres

,, 1 minute ,, ,, $\frac{60}{80}$,,

,, 60 minutes ,, ,, $\frac{60}{80}\times 60 = 45$ kilometres

PROBLEM 1.83. *A vehicle when climbing up a gradient, consumes petrol at a rate of one litre per eight kilometres, while climbing down it gives twelve kilometres per litre. Find its average rate of consumption for to and fro travel between two places situated at the two ends of a twenty five kilometres long gradient.*

SOLUTION. It is a problem of harmonic mean and not of a simple arithmetic mean.

$$\text{Harmonic mean}=\frac{2}{\frac{1}{8}+\frac{1}{12}}$$

$$=\frac{2}{\frac{5}{24}}=\frac{2}{1}\times\frac{24}{5}=9.6 \text{ Km per litre.}$$

Verification :

	Distance	Rate	Petrol consumed
up	25 Km.	8 K.P L.	$\frac{25}{8}$ litres
down	25 Km.	12. K. P. L.	$\frac{25}{12}$ litres
	50 Km.		$\frac{125}{24}$ litres

In $\frac{125}{24}$ litres he covers 50 Km.

,, 1 litre ,, ,, $\frac{50}{1}\times\frac{24}{125}$=9.6 Km. per litre

PROBLEM. 1.84. *An aeroplane flies along the four sides of a square at speeds of 100, 200, 300 and 400 kilometres per hour respectively. What is the average speed of the plane in its flight around the square.*

SOLUTION. In this problem harmonic mean is an appropriate average.

$$\text{H. M.}=\frac{4}{\frac{1}{100}+\frac{1}{200}+\frac{1}{300}+\frac{1}{400}}$$

$$=\frac{4}{(12+6+4+3)/1200}$$

$$=\frac{4}{\frac{25}{1200}}=\frac{4}{1}\times\frac{1200}{25}=192 \text{ Km. p.h.}$$

Verification :

Distance	Speed (Km. p. h.)	Time taken
100	100	60 minutes
100	200	30 minutes
100	300	20 minutes
100	400	15 minutes
400 Km.		125 minutes

In 125 minutes it covers 400 Kilometres

,, 1 minute ,, ,, $\frac{400}{125}$ Kilometres

,, 60 minutes ,, ,, $\frac{400}{125}\times 60=192$ Kilometres per hour.

PROBLEM 1.85. *A man having to drive 90 Kilometres wishes to achieve an average speed of 30 Kilometers per hour. For the first half of the journey he averages only 20 Km. p. h. What must be his average for the second half of the journey if his overall average is to be 30 Km. p. h.*

SOLUTION : In this problem harmonic mean would be more suitable.

Let x be the average speed for the second half of the journey

$$\text{Harmonic Mean} = \frac{n}{\frac{1}{a} \times \frac{1}{b}}$$

$$30 = \frac{2}{\frac{1}{20} + 1/x}$$

$$\text{or} \quad 2 = \frac{30}{20} + \frac{30}{x}$$

$$\text{or} \quad 2 = 1.5 + 30/x$$

$$\text{or} \quad .5 = 30/x$$

$$\text{or} \quad 5x = 30$$

$$\therefore \quad x = 60$$

Thus the average speed for the second half of the journey is 60 Kilometres per hour.

PROBLEM 1.86. *If A is arithmetic mean, H is harmonic mean and G is geometric mean of two numbres, show that (a) $AH = G^2$ and (b) $A > G > H$ (c) under what condition $A = G = H$. Illustrate by an example.*

SOLUTION : (*a*) Let a and b be the two numbers. Their A will be $(a+b)/2$ their G will be $\sqrt{a\,b}$ and their H will be

$$\text{Reciprocal } \frac{\frac{1}{a} + \frac{1}{b}}{2} \text{ or Reciprocal } \frac{\frac{b+a}{ab}}{\frac{2}{1}} \text{ or Reciprocal } \frac{a+b}{ab} \times \frac{1}{2}$$

$$\text{or Reciprocal } \frac{a+b}{2\,ab} \text{ or } \frac{2\,ab}{a+b}$$

$$\text{Now, } AH = \frac{a+b}{2} \times \frac{2\,ab}{a+b} = ab$$

$$\text{and} \quad ab = G^2 \quad \therefore \quad AH = G^2$$

(*b*) Let the numbers be 6 and 8

$$\text{Their } A \text{ will be} \quad \frac{6+8}{2} = 7$$

$$G = A.\,L.\left(\frac{0.7782 + 0.9031}{2}\right)$$

$$= A.\,L.\,\frac{1.6813}{2} = \text{A. L. } \cdot 8407$$

$$= 6.929$$

$$H \text{ will be Reciprocal } \left(\frac{0.1667 + 0.1250}{2}\right)$$

$$= \text{Reciprocal } \frac{.2917}{2} = \text{Reciprocal } .1458$$

$$\text{or} \quad H = 6.859$$

This satisfies the given condition.

(*c*) If all the given items are equal, $A = H = G$, and their value is same as given item.

Let the given items be 8,8.

$$A = \frac{8+8}{2} = 8$$

$$G = \text{A. L. } \left(\frac{\log 8 + \log 8}{2}\right)$$
$$= \text{A. L. } (\log 8)^2 = 8$$
$$H = \text{Reciprocal } \frac{\frac{1}{8}+\frac{1}{8}}{2}$$
$$= \text{Reciprocal } \frac{2\times\frac{1}{8}}{2} = 8$$

PROBLEM. 1.87. *If arithmetic mean and geometric mean of two value are 10 and 8 respectively, find the values.*

SOLUTION : Let two numbers be a and b

Then arithmetic mean $=(a+b)/2=10$ or $a+b=20$
Geometric mean $=\sqrt{a\times b}=8$ or $ab=64$
We know that $(a+b)^2-(a-b)^2=4\,ab$
By substituting the values, we get
$20^2-(a-b)^2=4\times 64$
$400-(a-b)^2=256$
$(a-b)^2=400-256=144$
Now $a-b=12$
and $a+b=20$
By adding, we get
$2a=20+12=32$
$\therefore$ $a=16$, and $b=4$
Hence the numbers are 16 and 4

REVISIONAL PROBLEM —1

PROBLEM. 1.88. *The following figures represent the number of book issued at the counter of a commerce college library on 12 different days :*

96, 180, 98, 75. 102, 270, 80, 100, 94, 75, 200, 600. Calculate the arithmet mean, median and mode for this data. Which of these would represent the abov data best?

(*Ans*: Mean=165, median=99, mode=75)
(Mean represents the data best)

PROBLEM 1.89. *The following table gives the age distribution of males i an Indian town. Find the mean age and the median age.*

Age group	*Males*	*Age group*	*Males*
0- 9	*2756*	*50-59*	*610*
10-19	*2124*	*60-69*	*245*
20-29	*1677*	*70-79*	*67*
30-39	*1481*	*80-89*	*6*
40-49	*1021*	*90-99*	*3*

(*Ans* : Mean=23.65, Median=20.22.)

PROBLEM 1.90. *A man purchases 5 rolls of ribbons each measuring 6 metres at the rate of 4, 6, 10, 12 and 15 meters a rupee respectively. Find the averag price of ribbons. Will it make any difference if the man purchases ribbons wort Rs. 4 from each of the above rolls ?*

(*Ans* : 9. 4 and 7.5)

PROBLEM 1.91. *Find the missing frequencies in the following distribution i N=100 and median of the distribution is 30.*

Marks	*No. of Students*
0-10	*10*
10-20	*?*
20-30	*25*
30-40	*30*
40-50	*?*
50-60	*10*

(*Ans* : 15 and 10)

PROBLEM 1.92. (*a*) *Explain clearly the relationship between Mean, Mode and Median in moderately skewed and normal distributions. Illustrate.*

(*b*) *Calculate the Value of the mode by the usual formula (after grouping if necessary)*

x	*f*
10- 20	*4*
20- 30	*6*
30- 40	*5*
40- 50	*10*
50- 60	*20*
60- 70	*22*
70- 80	*24*
80- 90	*6*
90-100	*2*
100-110	*1*

(*Ans* : $Z=65$)

PROBLEM 1.93. *Calculate Median and Mode from the following table :*

Income between *Rs.*	*No. of persons*
100 and 200	*15*
100 and 300	*33*
100 and 400	*63*
100 and 500	*83*
100 and 600	*100*

(*Ans* : Median=356.77, Mode 354.54)

PROBLEM 1.94. *Calculate Mean, Median and Mode from the fallowing data of the heights in centimeters of a group of students :*

161, 162, 163, 161, 163, 164, 164, 160, 165, 163, 164, 165, 166, 164

Now suppose that a group of students whose heights are 160, 166, 159, 168, 167 and 170 centimeters is added to the original group. Find Mean, Median and Mode of combined group.

(*Ans :* Mean=163.2 & 163.75
Median=163.5 & 164
Mode=164 & 164)

PROBLEM 1.95. *Calculate 3 Yearly moving average from the following data which relate to production and plot them along with the original data on a graph paper.*

Year	*1962*	*1963*	*1964*	*1965*	*1966*	*1967*	*1968*	*1969*	*1970*	*1971*
Production in lakh Rs.	*3*	*4*	*8*	*6*	*7*	*11*	*9*	*10*	*14*	*12*

(*Ans* : 5, 6, 7, 8, 9, 10, 11 and 12)

PROBLEM 1.96. *A train goes at a speed of 20 miles per hour for the first 16 miles, at a speed of 40 miles per hour for 20 miles. It covers the last ten miles at a speed of 15 miles per hour. Find out its average speed.*

(*Ans*: =*23.4 m. p. h.*)

PROBLEM 1.97. *The following are the monthly salaries in rupees of 30 employees of a firm.*

191	*239*	*226*	*219*	*200*	*187*	*165*	*177*	*199*	*195*
208	*227*	*186*	*248*	*216*	*176*	*169*	*188*	*212*	*218*
189	*216*	*197*	*205*	*195*	*180*	*186*	*206*	*193*	*235*

The firm gave bonuses as Rs. 10, 15, 20, 25, 30, 35, 40, 45 and 50 to employees in two respective salary groups : exceeding 160 but not exceeding 170, exceeding 170 but not exceeding 180 and so on up to exceeding 240 but not exceeding 250. Find the average bonus paid per employee.

(*Ans: Average bonus* =*Rs. 27.5*)

PROBLEM 1.98. *Fifty items sold in Dept. 'A' of the corner store had a mean price of 30 cents. 75 items sold in Dept. 'H' had a mean price of 20 cents. Find the mean price of commodities sold in departments 'A' and 'H'.*

(*Ans* : *Combined mean* =*24 cents*).

PROBLEM 1.99. *The mean annual salary paid to all employees of a company was Rs. 5000/— The mean annual salaries paid to male and female employees were Rs. 5200/— and Rs. 4200/— respectively. Determine the percentage of males and females employed by the company.* (*Males*=*80 per cent*; *Females*=*20 percent*)

PROBLEM 1.100. *Three persons A, B, and C can prepare a model in 20, 40 and 25 minutes respectively. What is their average speed ?*

(*Ans. one model in 23 minutes*)

PROBLEM 1.101. *You take a trip which entails travelling 900 Kilometers by train at an average speed of 60 Km. p. h. 300 Kilometers by boat at an average of 25 Km. p. h. 400 Kilometers by plane at 350 Km. p. h. and finally 15 Kilometers by taxi at 25 Km. p. h. What is your average speed for the entire distance (1615 Kilometers) ?*

Hence average speed for the entire distance=**31**.*6 Kilometers per hour.*

PROBLEM 1.102. *A person gets three successive annual rises in salary of 9%, 12% and 15% respectively, the percentage being reckoned in each case on his salary at the end of the previous year. What average percentage increase does this represent ?*

(*Ans* : *12.3%*)

PROBLEM 1.103. *There are three sections in B. Com. Ist year in a certain college. The number of students in each section and the average marks obtained by them in the paper statistics in the annual examination are as follows :—*

Section	*Average marks in Statistics*	*No. of Students*
A	*75*	*50*
B	*60*	*60*
C	*55*	*50*

Find the average marks obtained by the students of the sections taken together.

(*Ans.* =*63.125*)

PROBLEM 1.104. *From the following data calculate the crude accident rates of the two factories A and B and also the corrected accident data taking A as the standard. Comment upon your resvlt.*

Experience in Years	*Factory A*		*Factory B*	
	No of Workers	*No of Accidents*	*No of Workers*	*No of Accidents*
Under 5	*100*	*40*	*1000*	*300*
5— 15	*1500*	*150*	*500*	*40*
15— 25	*850*	*37*	*400*	*24*
25 and over	*50*	*8*	*100*	*6*
Total	*2500*	*235*	*2000*	*370*

(*C. A. R. Factory A*) =*94*)
(*C. A. R.* ,, *B*) =*135*)
(*C. A. R.* ,, *B*) =*81.6*)

PROBLEM 1.105. *Six types of workers are employed in each of two work shops but at different rates of wages as follows :*

Type of Workers	*Workshop A Rate of wages per worker Rs. paise*	*No of workers*	*Workshop B Rates of wages per worker Rs. paise*	*No of workers*
Mechanic	*4.50*	*2*	*5.00*	*18*
Fitter	*5.50*	*14*	*5.00*	*50*
Electrician	*6.00*	*20*	*6.25*	*8*
Carpenter	*5.00*	*7*	*5.50*	*12*
Smith	*5.00*	*6*	*5.50*	*10*
Clerk	*4.00*	*1*	*7.00*	*2*

In which of the two workshops is the average rate of wages per worker higher and by how much ? (Ans.: Factory A, 25 paise)

PROBLEM 1.106. *Compute crude and standardized death rates in the following and find out which of the populations is healthier.*

Age Group (Years)	Standard Population		Local Population	
	Population	Deaths	Population	Deaths
Under 5	6000	150	3,000	60
5-12	12,000	24	15,000	30
15-65	20,000	60	20,000	80
Above 65	8,000	320	8,000	400

(*Ans*. *G. D. R. of Standard Population* = 12·04
G. D. R. of Local Population = 12·31
S. D. R. of Local Population = 13·56)

PROBLEM 1.107. *Criticize the following arguments :*

(*i*) *The average depth of a river is 120 centimeters and the average height of the members of a family is 135 centimeters therefore, the family can surely cross the river.*

(*ii*) *The examination result of school A was 72% in a particular year. In the same year and at the same examination only 400 out of a total of 600 students were successful in school B. Therefore the teaching standard of the former school was decidedly better.*

(*iii*) *Death rate in the Navy is 6 per thousand and at Bombay the death rate is 26 per thousand. Therefore, it is better to be in the Navy than be employed in Bombay city.*

(*iv*) *An employer lowered the wages of an employee by 20% and later on raised by 50%. Hence there is no cause of complaint by the employee.*

Measures of Dispersion

The term 'dispersion' indicates an absolute or relative measure of the deviations of the items in a series from their average. It can be calculated in various ways. First we have dealt with problems in which Range has been calculated. Range is the difference between the biggest and smallest items of a series. We have then done some questions on Mean Deviation. Mean Deviation is the sum of deviations from an average. Mean Deviation is generally calculated from the Median. After this some problems of Quartile Deviation have been solved. Finally we have dealt with problems connected with the calculation of Standard Deviation which is the square root of the sum of the squares of the individual deviations from the Mean divided by the number of items.

RANGE

PROBLEM 2.1. *The following table gives the marks of students in two subjects :*

	English out of 100	*Optionals out of 400*
A	*20*	*145*
B	*40*	*200*
C	*90*	*322*
D	*30*	*150*
E	*80*	*275*
F	*75*	*285*

Find out the dispersion of the two series by Range method. Do you consider the results comparable ? If not, compute a relative measure of dispersion.

SOLUTION. Range of dispersion represent the difference between the values of the extremes, *i.e.*, the biggest and the smallest items of the data under review.

As such,

Range of dispersion of marks in English=90—20=70

and Range of dispersion of marks in Optionals=322—145=177

Range is an absolute measure of dispersion, and therefore, it is not possible to compare through it the relative dispersion of two series. If the absolute dispersions are reduced to relative ones, comparison is possible. This is done by dividing the range by the sum of the extreme items. The quotient called the co-efficient or ratio of dispersion.

Co-efficient of dispersion of marks in English$=\frac{70}{90+20}=\frac{70}{110}=.636$

and ,, ,, ,, Optionals$=\frac{177}{322+145}=\frac{177}{467}=.379$

Thus, the series of marks of English is more variable than the other series.

PROBLEM 2.2 *The following table gives the weights of one hundred persons. Compute the co-efficient of dispersion by the Method of Limits.*

Weight in kgs. of 100 persons

Class-interval	*No. of persons*
40-45	*4*
45-50	*13*
50-55	*8*
55-60	*14*
60-65	*9*
65-70	*16*
70-75	*17*
75-80	*9*
80-85	*8*
85-90	*2*
	100

SOLUTION. Computation of dispersion by the Method of Limits.

Weight in kgs. Class Interval	Mid-values (kgs.)	No. of persons
40-45	42.5	4
45-50	47.5	13
50-55	52.5	8
55-60	57.5	14
60-65	62.5	9
65-70	67.5	16
70-75	72.5	17
75-80	77.5	9
80-85	82.5	8
85-90	87.5	2

Range of dispersion $=90-40$ kgs. $=50$ kgs.

$$\text{Co-efficient of dispersion}=\frac{90-40}{90+40}=\frac{50}{130}=0.384$$

PROBLEM 2.3 *The following table gives the height of one hundred persons. Calculate the dispersion by Range Method.*

Height of 100 persons in centimetres

	Height in centimetres	*Frequency*
Below	*162*	*2*
,,	*163*	*8*
,,	*164*	*19*
,,	*165*	*32*
,,	*166*	*45*
,,	*167*	*58*
,,	*168*	*85*
,,	*169*	*93*
,,	*170*	*100*

SOLUTION. Computation of the dispersion of heights of 100 persons

Height in centimetres	Mid-values	Frequency
upto 162	161.5	2
162-163	162.5	6
163-164	163.5	11
164-165	164.5	13
165-166	165.5	13
166-167	166.5	13
167-168	167.5	27
168-169	168.5	8
169-170	169.5	7

Range of dispersion$=170-160=10$ centimetres

Co-efficient of dispersion$=\frac{170-160}{170+160}=\frac{10}{330}=.03$

QUARTILE DEVIATION

PROBLEM 2.4 *Calculate quartile deviation and its co-efficient of A' monthly earning for a year.*

Months	*Monthly earnings (Rs.)*
1	*239*
2	*250*
3	*251*
4	*251*
5	*257*
6	*258*
7	*260*
8	*261*
9	*262*
10	*262*
11	*273*
12	*275*

SOLUTION. Computation of quartile deviation of A's monthly earnings for a year.

Months	Monthly earnings Rs.
1	239
2	250
3	251
4	251
5	257
6	258
7	260
8	261
9	262
10	262
11	273
12	275

Q_1=the earnings of the $\frac{12+1}{4}$ *i.e.*, 3.25th month=Rs. 251.

Q_3=the earnings of the $\frac{3(12+1)}{4}$ *i.n.*, 9.75th month.

=Rs. 262.

Quartile Deviation or $Q.\ D.=\frac{Q_3-Q_1}{2}=$Rs. $\frac{262-251}{2}$

=Rs. 5.5

Quartile Co-efficient of Dispersion$=\frac{Q_3-Q_1}{Q_2+Q_1}=\frac{262-251}{262+251}$

$=\frac{11}{513}=.0213$

PROBLEM 2.5. *From the following table giving height of students calculate the semi-Interquartile Range and the co-efficient of Quartile Deviation.*

Height in centimetres	*No. of Students*
153	*25*
155	*21*
157	*28*
159	*20*
161	*18*
163	*24*
165	*22*
167	*18*
169	*23*

SOLUTION. Computation of the Semi-Interquartile Range.

Height in centimetres	No. of students	Cumulative Frequency
153	25	25
155	21	46
157	28	74
159	20	94
161	18	112
163	24	136
165	22	158
167	18	176
169	23	199

First Quartile or,

Q_1 = the height of the $\frac{199+1}{4}$ *i.e.*, 50th student,

= 157 centimetres

Third Quartile or

Q_3 = the height of the $\frac{3(199+1)}{4}$ *i.e.*, 150th student.

= 165 centimetres

Semi-interquartile Range or

$$\text{the Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{165-157}{2} = 4 \text{ centimetres}$$

$$\text{Quartile Co-efficient of Dispersion} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{165-157}{165+157} = .025.$$

PROBLEM 2.6 *Calculate the Semi-Interquartile Range and its co-efficient of the marks of 59 students in Economics as given below:—*

Marks group	*No. of Students*
0—10	*4*
10—20	*8*
20—30	*11*
30—40	*15*
40—50	*12*
50—60	*6*
60—70	*3*

SOLUTION. Computation of the Semi-Interquartile Range

Marks group	No. of students	Cumulative Frequency
0—10	4	4
10—20	8	12
20—30	11	23
30—40	15	38
40—50	12	50
50—60	6	56
60—70	3	59

First Quartile or

Q_1=the marks of the $\frac{59}{4}$ *i. e.*, 14.75th student which lies in the 20—30 marks group.

By interpolation.

$$=20+\frac{30-20}{11}(14.75-12)=22.5 \text{ marks}$$

Third Quartile or

Q_3=the marks of the $\frac{3(59)}{4}$ *i.e.*., 44.25th student which lies in the 40—50 marks group.

By interpolation,

$$=40+\frac{50-40}{12}(44.25-38)=54.2 \text{ marks approx.}$$

$$\text{Semi-Interquartile Range}=\frac{Q_3-Q_1}{2}=\frac{45.2-22,5}{2}=11.35 \text{ marks}$$

$$\text{Co-efficient of the S. I. Range}=\frac{Q_3-Q_1}{Q_3+Q_1}=\frac{45.2-22.5}{45.2+22.5}=.338.$$

PROBLEM 2.7 *From the following table compute the quartile deviation :*

Size	*4-8*	*8-12*	*12-16*	*16-20*	*20-24*	*24-28*	*28-32*	*32-36*	*36-40*
Frequency	*6*	*10*	*18*	*30*	*15*	*12*	*10*	*6*	*2*

SOLUTION. Computation of the quartile deviation.

Size	Frequency	Cumulative Frequency
4—8	6	6
8—12	10	16
12—16	18	34
16—20	30	64
20—24	15	79
24—28	12	91
28—32	10	101
32—36	6	107
36—40	2	109

First quartile or

Q_1 =the size of the $\frac{109}{4}$ *i. e.*, 27.25th item which lies in the 12—16 group

By interpolation ,

$$=12+\frac{4}{18}(27.25-16)=14.5$$

Third Quartile

Q_3=the size of the $\frac{3(109)}{4}$ *i.e.*, 81.75th, item which lies in the 24—28 group.

By interpolation,

$$=54+\frac{28-24}{12}(81.75-79)=24.92$$

$$\text{Quartile Deviation}=\frac{Q_3-Q_1}{2}=\frac{24.92-14.50}{2}=\frac{10.42}{2.}=5.21.$$

AVERAGE DEVIATION

PROBLEM 2.8 *The following are the marks obtained by a batch of 9 students in a certain test :—*

Serial Number	*Marks (out of 100)*
1	*68*
2	*49*
3	*32*
4	*21*
5	*54*
6	*38*
7	*59*
8	*66*
9	*41*

Calculate the Mean Deviation of the series.

SOLUTION. Calculation of Mean Deviation of the series of marks of 9 students (arranged in ascending order of magnitude).

Students	Marks (*m*)	Deviations from median (49) (+&—signs ignored) (*dm*)
1	21	28
2	32	17
3	38	11
4	41	8
5	49	0
6	54	5
7	59	10
8	66	17
9	68	19
		$\Sigma dm=115$

Median $=$ Value of $\frac{n+1}{2}$ items

$=$Value of 5th item$=$49 marks

Mean Deviation $\delta m = \frac{\Sigma dm}{n}$

Where Σdm represents the summation of the deviations from the median ; and n, the number of items,

$\delta m = \frac{115}{9}$ marks.$=$12.8 marks.

Alternative Method

Marks arranged in ascending order of magnitude.

Marks (*m*)
21
32
38
41
49
54
59
66
68

Sum of items $<$ median

$= (21+32+38+41) = 132$ (mx)

Sum of items $>$ median

$= (54+59+66+68) = 247)$ (my)

Mean Deviation $= \frac{1}{n}(my - mx) = \frac{1}{9}$ $(247 - 132)$

$= \frac{1}{9} \times 115 = 12.8$ Marks. approx.

PROBLEM 2.9 *The following are the daily rents of 18 houses in a certai locality :—*

Rs.	*P.*	*Rs.*	*P.*
6	*50*	*6*	*25*
5	*00*	*3*	*00*
5	*25*	*9*	*00*
5	*50*	*5*	*00*
5	*25*	*4*	*00*
4	*75*	*3*	*00*
4	*00*	*3*	*75*
5	*00*	*4*	*00*
4	*50*	*3*	*00*

Calculate the Mean Deviation of this group.

SOLUTION. Calculation of mean Deviation of the series of daily rents o 18 houses (arranged in ascending order of magnitude).

Houses	Monthly Rent Rs. (m)	Deviations from median (Rs. 4.875) (+& – signs ignored) (dm)
1	3.00	1.875
2	3.00	1.875
3	3.00	1.875
4	3.75	1.125
5	4.00	.875
6	4.00	.875
7	4.00	.375
8	4.50	.375
9	4.75	.125
10	5.00	.125
11	5.00	.125
12	5.00	.125
13	5.25	.375
14	5.25	.375
15	5.50	.625
16	6.25	1.375
17	6.50	1.625
18	9.00	4.125
		$\Sigma dm = 18.250$

Median = Value of $\frac{n+1}{2}$ items.

= Value of 9.5th item. = Rs, 4.875

Mean Deviation from the Median, or, $\delta m = \frac{\Sigma dm}{n}$;

Where Σdm represents the summation of the deviations from the median and n, the number of items,

$$\delta m = \text{Rs. } \frac{18.250}{18} = \text{Rs. } 1.014$$

Thus the mean deviation of the rents of 18 houses of the said locality is Rs. 1.014.

Alternative method

Sum of items <median or $mx = 34.50$
Sum of items >median or $my = 52.75$

$$\text{Mean Deviation} = \frac{1}{n}(my - mx) = \frac{1}{18}(52.75 - 34.50)$$

$$= \frac{1}{18} \times 18.25 = 1.014 \text{ rupees.}$$

PROBLEM 2.10 *Calculate the mean deviation (from the arithmetic average) and its co-efficient for the following prices of a Government Security.*

Rs. 100 $\frac{1}{2}$, $\frac{1}{4}$, $\frac{3}{4}$, $\frac{5}{8}$, $\frac{3}{4}$, $\frac{1}{8}$, $\frac{6}{16}$, $\frac{10}{16}$, $\frac{8}{16}$, $\frac{1}{8}$.

SOLUTION. Calculation of mean deviation from the arithmetic average.

Prices (m)	Deviation from arithmetic average (100.425) (+ & − signs ignored) (d)
100.500	.075
100.250	.175
100.375	.050
100.625	.200
100.750	.325
100.125	.300
100.375	.050
100.625	.200
100.500	.075
100.125	.300
$\Sigma m = 1004.250$	$\Sigma d = 1.750$

$$\text{Arithmetic average} = \frac{\Sigma m}{n} = \frac{1004.250}{10} = \text{Rs. } 100.425$$

$$\text{Mean Deviation from the Arith. Av., or } \delta = \frac{\Sigma d}{n}$$

$$= \text{Rs. } \frac{1.750}{10} = \text{Re. } .175$$

$$\text{Co-efficient of Dispersion from the Arith. av.} = \frac{\delta}{a}$$

$$= \frac{.175}{100.425} = \cdot 0017 \text{ approx.}$$

Thus, the mean deviation of the given prices of Government Security is .175 and its co-efficient is .0017 approx.

Alternate Method

Prices arranged in ascending order

Prices
100.125
100.125
100.250
100.375
100.375
100.500
100.500
100.625
100.625
100.750
1004.250

Arithmetic average or $a = \frac{1004.250}{10} = 100.425$

Number of items smaller than arithmetic average or $nx = 5$ and their total or $ay = 501.250$

Number of items bigger than arithmetic average or $ny =$ and their total or $ay = 503.000$

Mean Deviation $= \frac{1}{n}[(ay - ny \times a) + (nx \times a - ax)]$

$$= \frac{1}{10}[(503 - 5 \times 100.425) + (5 \times 100.425 - 501.25)]$$

$$= \frac{1}{10}(.875 + .875) = \frac{1.750}{10} = .175 \text{ rupee.}$$

PROBLEM 2.11 *Find Mean Deviation of the distribution given below :—*

No. of accidents	*Persons having said number of accidents*
0	*15*
1	*16*
2	*21*
3	*10*
4	*17*
5	*8*
6	*4*
7	*2*
8	*1*
9	*2*
10	*2*
11	—
12	*2*
Total	*100*

SOLUTION. Calculation of the Mean Deviation.

No. of accidents (*m*)	Persons having said number of accidents (*f*)	Deviation from Median (2) (+ & – signs ignored) (*dm*)	Total Deviations (*fdm*)
0	15	2	30
1	16	1	16
2	21	0	0
3	10	1	10
4	17	2	34
5	8	3	24
6	4	4	16
7	2	5	10
8	1	6	6
9	2	7	14
10	2	8	16
11	—	9	0
12	2	10	20
	$n = 100$		$\Sigma dm = 196$

Median no. of accidents, = the no. of accidents, the $\frac{100+1}{2}$ or

the 50·5th person had.

$=2$

Mean Deviation $=\frac{\Sigma fdm}{n}=\frac{196}{100}=1.96$ accidents.

PROBLEM 2.12 *Compute mean deviation from arithmetic mean from the following data :*

Height (centimetres)	*158*	*159*	*160*	*161*	*162*	*163*	*164*	*165*	*166*
No. of students	*15*	*20*	*32*	*35*	*33*	*22*	*20*	*10*	*8*

SOLUTION. Calculation of mean deviation from arithmetic average.

Height	No. of students	Deviations from the as. av. 162	Total deviations fdx	(d′) + and − signs ignored	Total deviations + and − signs ignored
158	15	−4	−60	3.51	52.65
159	20	−3	−60	2.51	50.20
160	32	−2	−64	1.51	48.32
161	35	−1	−35	.51	17.85
162	33	0	0	.49	16.17
163	22	+1	+22	1.49	32.78
164	20	+2	+40	2.49	49.80
165	10	+3	+30	3.49	34.90
166	8	+4	+32	4.49	35.92
	$n=195$		$\Sigma fdx=-95$		$\Sigma fd=338.59$

Actual arithmetic average $=162-\frac{95}{195}=161.51$ approx.

Mean deviation $(\delta_a)=\frac{\Sigma fd'}{n}=\frac{338.59}{195}=1.74$

PROBLEM 2.13 *Calculate mean deviation from median from the following data :*

	No. of students
Marks less than 80	*100*
,, ,, *70*	*90*
,, ,, *60*	*80*
,, ,, *50*	*60*
,, ,, *40*	*32*
,, ,, *30*	*20*
,, ,, *20*	*13*
,, ,, *10*	*5*

SOLUTION.

Marks (*m*)	Mid-values (*mv*)	No. of students (*f*)	Cumulative frequency	Deviation from median 46.43 (+and − signs ignored) *dm*	Total deviations (*fdm*)
0-10	5	5	5	41.43	207.15
10-20	15	8	13	31.43	251.44
20-30	25	7	20	21.43	150.01
30-40	35	12	32	11.43	137.16
40-50	45	28	60	1.43	40.04
50-60	55	20	80	8.57	171.40
60-70	65	10	90	18,57	185,70
70-80	75	10	100	28.57	285.70
					Σfdm = 1428.60

Median = size of $\frac{n}{2}$th item *i.e.*, $\frac{100}{2}$ = 50th item which lies in 40-50 group.

By interpolation,

$$M=40+\frac{50-40}{28}(50-32)=46.43$$

Mean deviation from median $=\frac{\Sigma fdm}{n}=\frac{1428.60}{100}=14.286.$

PROBLEM 2.14 *Calculate the mean deviation (from mean) of the following series :—*

Marks	*No. of Students*
0-10	*5*
10-20	*8*
20-30	*15*
30-40	*16*
40-50	*6*

SOLUTION. Calculation of mean deviation

Marks	Mid-values	Step Deviation from as. av. (25)		No. of students		Deviation from actual average (27)	
	(*mv*)	(*d*)	(*d′*)	(*f*)	(*fd′*)	(*d*)	(*fd*)
0—10	5	—20	—2	5	—10	22	110
10—20	15	—10	—1	8	— 8	12	96
20—30	25	0	0	15	0	2	30
30—40	35	+10	+1	16	+16	8	128
40—50	45	+20	+2	6	+12	18	108
				$\Sigma f=50$	$\Sigma fd'=+10$		$\Sigma fd=472$

Arithmetic average or $a=25+\left(\frac{10}{50}\times 10\right)=27$

Mean deviation $=\frac{\Sigma fd}{\Sigma f}=\frac{472}{50}$ marks=9.44 marks.

Alternate Method.

Marks	Mid-values (*mv*)	Step deviations from as. av. (25) (*d'*)	Number of Students (*f*)	(*fd'*) (+& - signs ignored)	(*fd'*) (with + & — signs)
0—10	5	—2	5	10	—10
10—20	15	—1	8	8	— 8
20—30	25	0	15	0	0
30—40	35	+1	16	16	+16
40—50	45	+2	6	12	+12
			$\Sigma f=50$	$\Sigma fd=46$	$\Sigma fd'=+10$

Actual arithmetic average $=25+\left(\frac{10}{50}\times 10\right)=27.$

Total deviations from assumed average of $25=46\times 10=460$

Adjustments

Number of items less than the actual arithmetic average (27)=28

Number of items more than the actual arithmetic average (27)=22

Difference between actual and assumed average=2.

Total deviations from actual average or (Σfd), where actual and assumed averages are in the same class-interval.

$$=460+(28\times 2)-(22\times 2)$$
$$=460+56-44=472$$

$$\text{Mean Deviation}=\frac{\Sigma fd}{n}$$

$$=\frac{472}{50}=9.44 \text{ marks.}$$

PROBLEM 2.15 *Calculate the mean deviation from the following data. What light does it throw on the social conditions of the community?*

Difference in age between husband and wife in a particular community.

Difference	*Frequency*
0— 5	*449*
5—10	*705*
10—15	*507*
15—20	*281*
20—25	*109*
25—30	*52*
30—35	*16*
35—40	*4*

SOLUTION. Calculation of Mean Deviation

Difference in years	Mid-values	Frequency	Deviation from the as. av. (12,5)		Total deviations from the average assumed	Deviation from the av.(10.5) +&— signs ignored	Total deviations
(m)	(mv)	(f)	(dx)	dx/i where $i=5$	(fdx)	(d)	(d)
0— 5	2.5	449	−10	−2	−898	8	3592
5—10	7.4	705	− 5	−1	−705	3	2115
10—15	12.5	507	0	0	0	2	1014
15—20	17.5	281	+ 5	+1	+281	7	1967
20—25	22.5	109	+10	+2	+218	12	1308
25—30	27.5	52	+15	+3	+156	17	884
30—35	32.5	16	+20	+4	+ 64	22	352
35—40	37.5	4	+25	+5	+ 20	27	108
		$n=2123$			$\Sigma dx=-864$		$\Sigma fd=11340$

$$\text{Arithmetic average}=x+\frac{\Sigma fdx}{n}\times i$$

$$=12.5+\left(\frac{-864}{2123}\times 5\right)=10.47$$

$$\text{Mean Deviation}=\frac{\Sigma fd}{n}=\frac{11340}{2123}=5.3 \text{ approx.}$$

Alternate Method

Total Deviation from assumed average (±signs ignored)

$=(2342\times 5)=11710$

Number of items smaller than actual arithmetic average (10.5)

$=1154$

Number of items bigger than actual arithmetic average (10.5)

$=969$

Difference between actual and assumed average

$=-2$

Total deviations from actual arithmetic average (where actual and assumed average are in the same class-interval)

$$=11710+[(1154\times -2)-(969\times -2)]$$

$$=11710-2308+1938=11340$$

$$\text{Mean Deviation}=\frac{11340}{2123}=5.3 \text{ approx.}$$

PROBLEM 2.16 *The following table gives the age distributions of students admitted to a college in the years 1964 and 1968. Find which of the two groups is more variable in age.*

Age	*Number of students admitted in* *1964*	*1968*
15—	—	*1*
16—	*1*	*6*
17—	*3*	*34*
18—	*8*	*22*
19—	*12*	*35*
20—	*14*	*20*
21—	*14*	*7*
22—	*5*	*19*
23—	*2*	*3*
24—	*3*	—
25—	*1*	—
26—	—	*1*
27—	*1*	—
Total	*64*	*148*

SOLUTION. Computation of the Mean Deviation of the 1964 group.

Age group (*m*)	Mid-value of the group (*mv*)	Frequency (*f*)	Deviation from the as. av. (19.5) (*dx*)	Total deviation from the as. av. (*fdx*)	Deviation from the average (20.7) (+&— signs ignored) (*d*)	Total deviation from the average (*fd*)
15—16	15.5	0	−4	0	5.2	0
16—17	16.5	1	−3	−3	4.2	4.2
17—18	17.5	3	−2	−6	3.2	9.6
18—19	18.5	8	−1	−8	2.2	17.6
19—20	19.5	12	0	0	1.2	14.4
20—21	20.5	14	+1	+14	.2	2.8
21—22	21.5	14	+2	+28	.8	11.2
22—23	22.5	5	+3	+15	1.8	9.0
23—24	23.5	2	+4	+ 8	2.8	5.6
24—25	24.5	3	+5	+15	3.8	11.4
25—26	25.5	1	+6	+ 6	4.8	4.8
26—27	26.5	—	+7	0	5.8	0
27—28	27.5	1	+8	+ 8	6.8	6.8
		$n=64$		$\Sigma dx=+77$		$\Sigma d=97.4$

Arithmetic average or $a=x+\frac{\Sigma fdx}{n}=19.5+\frac{77}{64}=20.7$ years.

Mean Deviation or $\delta=\frac{\Sigma fd}{n}=\frac{97.4}{64}=1.52$ years approx.

Mean Coefficient of Dispersion$=\frac{\delta}{a}=\frac{1.52}{20.7}=.0734$

Alternate Method

Total deviation from assumed average ± signs ignored=111.

Note. Where actual and assumed averages are in different class intervals a special adjustment is necessary. In such cases the frequency of the class in which the actual mean lies is treated separately. It is multiplied by the difference of th deviations of the mid-value from the actual and assumed average. The produc so obtained is subtracted from the total deviations from the assumed mean.

Thus :—

Number of items smaller than the actual arithmetic average (20.7)=24 (fre quency of mean class being ignored).

Number of items bigger than actual arithmetic average (20.7)=26

Frequency of the mean class=14.

Difference between actual and assumed averages=1.2

Difference of deviations of mid-value of mean class from the actual an assumed averages$=(20.5-20.7)-(20.5-19.5)$

$=-.2-1=.8$(+ signs ignored)

Total deviations from actual mean

$=111+(24\times1.2)-(26\times1.2)-(14\times.8)$

$=111+28.8-31.2-11.2=97.4$

Mean deviation $=\dfrac{97.4}{64}=1.52$

Mean Co-efficient of Dispersion $=\dfrac{1.52}{20.7}=.0734$

Computation of the Mean Deviation of the 1968 group

Age group (m)	Mid-values of the group ($m.v.$)	Fre-quency (f)	Devia-tion from the as. av. (19.5) (dx)	Total Deviation from the as. av. (fdx)	Deviation from the average (19.5) + & − signs ignored (d)	Total Deviation from the average (fd)
15—16	15.5	1	−4	−4	4	4
16—17	16.5	6	−3	−18	3	18
17—18	17.5	34	−2	−68	2	68
18—19	18.5	22	−1	−22	1	22
19—20	19.5	35	0	0	0	0
20—21	20.5	20	+1	+20	1	20
21—22	21.5	7	+2	+14	2	14
22—23	22.5	19	+3	+57	3	57
23—24	23.5	3	+4	+12	4	12
24—25	24.5	—	+5	—	5	—
25—26	25.5	—	+6	—	6	—
26—27	26.5	1	+7	+7	7	7
27—28	27.5	—	+8	—	8	—
		$n=148$		$\Sigma fdx=-2$		$\Sigma fdx=22$

Arithmetic average or $a=\dfrac{\Sigma fdx}{n}=19.5+\dfrac{-2}{148}=19.49$

years appro

Mean Deviation or $\delta=\dfrac{\Sigma fd}{n}=\dfrac{222}{148}=1.5$ years appro

Mean Coefficient of Dispersion $=\dfrac{\delta}{a}=\dfrac{1.5}{19.5}=.077$

On comparison of the mean coefficients of dispersion, it is obvious that 1968 group is more variable than the 1964 group.

PROBLEM 2.17 *(a) Calculate median coefficient of dispersion (b) mean coefficient of dispersion and (c) mode coefficient of dispersion from the following data and show that mean deviation about the median is the least,*

Size of Items	*4*	*6*	*8*	*10*	*12*	*14*	*16*
Frequency	*2*	*4*	*5*	*3*	*2*	*1*	*4*

SOLUTION :

Size of Items (m)	Frequency (f)	Cumulative frequency (df)	Deviations from 8 (dx)	Total deviations (fdx)
4	2	2	-4	-8
6	4	6	-2	-8
8	5	11	0	0
10	3	14	$+2$	$+6$
12	2	16	$+4$	$+8$
14	1	17	$+6$	$+6$
16	4	21	$+8$	$+32$
	$n=21$			$\Sigma fdx=36$ $\Sigma \lvert fdx \rvert=68$

$$\text{Arithmetic mean}=8+\frac{36}{21}=9.71$$

$$\text{Median Item}=\left(\frac{n+1}{2}\right)\text{th item i.e., }\frac{21+1}{2}=11\text{th item}$$

Size of 11th item is 8 ∴ Median is 8.

Mode : If we group the frequencies we shall find that frequencies 5 will determine the mode. This is clear from the following test also.

Size :	6	8	16
	2	4	1
	4	5	4
	5	3	..
	11	12	5

Thus the mode also is 8.

Mean deviation about the mean :

Total deviation from assumed mean=68

No. of items smaller than actual mean=11

No. of items bigger than actual mean=10

Difference between actual and assumed average=1.71

Total deviations from actual mean $=68+(11-10)\times1.71=69.71$

$$\text{Mean Deviation}=\frac{69.71}{21}=3.3195$$

Mean deviation about the Median and Mode (as in this problem both are the same).

$$\delta z \text{ or } \delta m=\frac{68}{21}=3.238 \text{ approx.}$$

From the above it is clear that Mean deviation about the Median is the least.

STANDARD DEVIATION

PROBLEM 2.18 *Find out the Standard Deviation of the following items :-*

8, 10, 12, 14, 16, 18, 20, 22, 24, 26

SOLUTION. Calculation of Standard Deviation.

Size of items (m)	Deviation from mean (17) (d)	Deviation square (d^2)
8	−9	81
10	−7	49
12	−5	25
14	−3	9
16	−1	1
18	+1	1
20	+3	9
22	+5	25
24	+7	49
26	+9	81
$\Sigma m = 170$		$\Sigma d^2 = 330$

Arithmetic Average or $a = \frac{\Sigma m}{n} = \frac{170}{10} = 17$

Standard Deviation or $\sigma = \sqrt{\frac{\Sigma d^2}{n}} = \sqrt{\frac{330}{10}} = 5{\cdot}745$

Alternate Method

Size of items (m)	(m^2)
8	64
10	100
12	144
14	196
16	256
18	324
20	400
22	484
24	576
26	676
$\Sigma m = 170$	$\Sigma m^2 = 3220$

Standard Deviation or σ

$$= \sqrt{\frac{\Sigma m^2 - (\Sigma m)^2/n}{n}}$$

$$= \sqrt{\frac{3220 - (170)^2/10}{10}}$$

$$= \sqrt{\frac{3220 - 28900/10}{10}}$$

$$= \sqrt{\frac{3220 - 2890}{10}}$$

$$=\sqrt{\frac{330}{10}}$$

$$= 5{\cdot}745$$

PROBLEM 2.19. *Find out the Standard Deviation of the height of 10 men given below :—*

160, 160, 161, 162, 163, 163, 163, 164, 164, 170

SOLUTION. Calculation of Standard Deviation of the height.

Height in cms. (m)	Deviation from mean (163) (d)	Deviation square (d^2)
160	—3	9
160	—3	9
161	—2	4
162	—1	1
163	0	0
163	0	0
163	0	0
164	+1	1
164	+1	1
170	+7	49
$\Sigma m = 1630$		$\Sigma d^2 = 74$

Arithmetic average or $a=\frac{\Sigma m}{n}=\frac{1630}{10}=163$

Standard Deviation or $\sigma=\sqrt{\frac{\Sigma d^2}{n}}$

$$=\sqrt{\frac{74}{10}}$$

$$=\sqrt{7.4}$$

$$= 2.72$$

Alternate Method

Size of item (m)	(m^2)
160	25,600
160	25,600
161	25,921
162	26,244
163	26,569
163	26,569
163	26,569
164	26,896
164	26,896
170	28,900
$\Sigma m = 1630$	$\Sigma m^2 = 2{,}65{,}764$

Standard Deviation or $\sigma=\sqrt{\frac{\Sigma m^2=(\Sigma m)^2/n}{n}}$

$$=\sqrt{\frac{2{,}65{,}764-(1630)^2/10}{10}}$$

$$=\sqrt{\frac{2{,}65{,}764-26{,}56{,}900/10}{10}}$$

$$=\sqrt{\frac{74}{10}}$$

$$=2.72$$

PROBLEM 2.20. *Nine students of B. Com. class of a college have obtaine the following marks in statistics out of 100 marks. Calculate the standard deviatio of marks obtained.*

S. No.	*Marks*
1	*5*
2	*10*
3	*20*
4	*25*
5	*40*
6	*45*
7	*48*
8	*70*
9	*80*

SOLUTION. Calculation of Standard Deviation

Size of items (m)	Deviations from assumed mean 40 (dx)	Square of deviations (dx^2)
5	−35	1225
10	−30	900
20	−20	400
25	−15	225
40	0	0
45	+5	25
48	+8	64
70	+30	900
80	+40	1600
$n=10$	$\Sigma dx=-17$	$\Sigma dx^2=5339$

Standard Deviation or $\sigma=\sqrt{\frac{\Sigma dx^2}{n}-\left(\frac{\Sigma dx}{n}\right)^2}$

$$=\sqrt{\frac{5339}{10}-\left(\frac{-17}{n}\right)^2}$$

$$=\sqrt{533.9-2.89}$$

$$=\sqrt{531.01}$$

$$=\ 23.05 \text{ approx.}$$

PROBLEM 2.21. *Calculate the Deviation of the following data with rega to 2,298 families in the U. K.*

Number of persons in the family	*Number of families*
1	*165*
2	*552*
3	*580*
4	*433*
5	*268*
6	*148*
7	*77*
8	*41*
9	*20*
10	*8*
11	*5*
12	*1*
Total	*2,298*

SOLUTION. Calculation of Standard Deviation.

Size of family (m)	No. of families (f)	Deviation from the assumed average (4) (dx)	Total Deviation (fdx)	Square of Deviation (dx^2)	Frequency × Square of Deviation (fdx^2)
1	165	−3	−495	9	1485
2	552	−2	−1104	4	2208
3	580	−1	−580	1	580
4	433	0	0	0	0
5	268	+1	+268	1	268
6	148	+2	+296	4	592
7	77	+3	+231	9	693
8	41	+4	+164	16	656
9	20	+5	+100	25	500
10	8	+6	+ 48	36	288
11	5	+7	+ 35	49	245
12	2	+8	+ 8	64	64
	$n = 2298$		Σfdx—1029		$\Sigma fdx^2 = 7579$

$$(a-x) = \frac{-1029}{2298} = -\cdot 45$$

Standard Deviation of the size of family $= \sqrt{\dfrac{\Sigma fdx^2 - n\,(a-x)^2}{n}}$

$$= \sqrt{\frac{7579 - 2298(-.45)^2}{2298}}$$

$$= \sqrt{3.14}$$

$$= 1.77$$

PROBLEM 2.22. *Find out the mean daily earnings and standard deviation of earnings from the following data :—*

	Rs.	
36 men get at the rate of	*5.0*	*per man per day.*
40 men get at the rate of	*5.5*	*per man per day.*
90 men get at the rate of	*6.0*	*per man per day.*
138 men get at the rate of	*6.5*	*per man per day.*
80 men get at the rate of	*7.0*	*per man per day.*
61 men get at the rate of	*7.5*	*per man per day.*
25 men get at the rate of	*8.0*	*per man per day.*

SOLUTION. Calculation of the Mean and Standard Deviation of earnings.

Earnings (Rs.) (m)	Number of persons (f)	Deviation from as. av. (6.5) (d')	(d'^2)	fd'	fd'^2
5.0	36	−1.5	2.25	−54.0	81.00
5.5	40	−1.0	1.00	−40.0	40.00
6.0	90	−0.5	2.25	−45.0	22.50
6.5	138	0	0	0	0
7.0	80	+0.5	0.25	+40.0	20.00
7.5	61	+1.0	1.00	+61.0	61.00
8.0	25	+1.5	2.25	+37.5	56.25
	$\Sigma f = 470$			$\Sigma fd' = -0.5$	$\Sigma fd'^2 = 280.75$

$$\text{Mean or } a = x + \left(\frac{\Sigma fd'}{a}\right)$$

$$= 6.5 + \left(\frac{-.5}{470}\right) = 6.499 \text{ rupees.}$$

$$\text{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma fd'^2}{n} - \left(\frac{\Sigma fd'}{n}\right)^2}$$

$$= \sqrt{\frac{280\cdot75}{470} - \left(\frac{-\cdot5}{470}\right)^2}$$

$$= \sqrt{\frac{2\,80.75}{470} - (\cdot001)}$$

$$= \sqrt{.596} \text{ or } \cdot77 \text{ rupees.}$$

PROBLEM 2.23. *Calculate the standard deviation for the following table giving the distribution of 542 members of the House of Commons.*

Age	No. of members.
20—	3
30—	61
40—	132
50—	153
60—	140
70—	51
80—	2
Total	542

SOLUTION. Calculation of the Standard Deviation of the age distribution of 542 members of the House of Commons

Age group (*m*)	Mid-value (*mv*)	Frequency (*f*)	Deviation from the assumed av. (55) dx	Total Deviation fdx	Square of Deviations dx^2	Frequency × Square of Deviations fdx^2
20—30	25	3	—30	— 90	900	2700
30—40	35	61	—20	— 1220	400	24400
40—50	45	132	—10	—1320	100	13200
50—60	55	153	0	0	0	0
60—70	65	140	+10	+1400	100	14000
70—80	75	51	+20	+1020	400	20400
80—90	85	2	+30	+ 60	900	1800
		$n=542$		$\Sigma fdx=$ -150		$\Sigma fdx^2=76500$

$$(a-x)=\frac{-150}{542}=-.276$$

$$\text{Standard Deviation}=\sqrt{\frac{\Sigma fdx^2-n\,(a-x)^2}{n}}$$

$$=\sqrt{\frac{76500-542(.276)^2}{542}}$$

$$=\sqrt{141.7}$$

$$=11.9$$

The following method will also give us the same result.

$$\text{Standard Deviation of the age distribution}=\sqrt{\frac{fdx^2}{n}-\left(\frac{fdx}{n}\right)^2}$$

$$=\sqrt{\frac{76500}{542}-\left(\frac{-150}{542}\right)^2}$$

$$=\sqrt{141.07}$$

$$=11.9 \text{ years.}$$

PROBLEM 2.24. *Explain the method of computing the Standard deviation of a frequency distribution from a working origin different from the arithmetic mean.*

Calculate the Standard Deviation for the data given below using the interval 50—59, as working origin :—

Class-interval	*Frequency*	*Class-interval*	*Frequency*
0— 9	*2*	*80 – 89*	*12*
10—19	*4*	*90— 99*	*9*
20—29	*23*	*100—109*	*6*
30—39	*30*	*110—119*	*10*
40—49	*40*	*120—129*	*3*
50—59	*45*	*130—139*	*1*
60—69	*35*	*140—149*	*1*
70—79	*25*	*150—159*	*3*
Total			*249*

SOLUTION. Calculation of standard deviation, 50—59 interval as working origin.

Class interval (m)	Mid value ($m.v.$)	Fre-quency (f)	Devia-tion from the as. av. (54.4) (dx)	Total Devia-tions (fdx)	Square of Devia-tions (dx^2)	Frequency × Square of Deviation (fdx^2)
0— 9	4.5	2	— 50	—100	2500	5000
10— 19	14.5	4	— 40	—160	1600	6400
20— 29	24.5	23	— 30	—690	900	20700
30— 39	34,5	30	— 20	—600	400	12000
40— 49	44.5	40	— 10	—400	100	4000
50— 59	54.5	45	0	0	0	0
60— 69	64.5	35	+ 10	+350	100	3500
70— 79	74.5	25	+ 20	+500	400	10000
80— 89	84.5	12	+ 30	+360	900	10800
90— 99	94.5	9	+ 40	+360	1600	14400
100—109	104.5	6	+ 50	+300	2500	15000
110—119	114.5	10	+ 60	+600	3600	36000
120—129	124.5	3	+ 70	+210	4900	14700
130—139	134.5	1	+ 80	+ 80	6400	6400
140—149	144.5	1	+ 90	+ 90	8100	8100
150—159	154.5	3	+100	+300	10000	30000
		n=249		$\Sigma fdx=120$		Σfdx^2=1,97,000

$$(a-x)=\frac{1200}{249}=4.8$$

Standard Deviation or $\sigma=\sqrt{\dfrac{\Sigma fdx^2-n\,(a-x)^2}{n}}$

$$=\sqrt{\frac{197000-249\,(4.8)^2}{249}}$$

$$=\sqrt{768.125}$$

$$=27.71$$

PROBLEM. 2.25. *The following table gives the frequency distribution of expenditure on food per family per month among working class families in two localities. Find the arithmetic average and the Standard Deviation of the expenditure at both places.*

Range of expenditure in Rs. per month		*No. of families*	
		Place A	*Place B*
Rs.	*203—206*	*28*	*39*
,,	*206—209*	*292*	*284*
,,	*209—212*	*389*	*401*
,,	*212—215*	*212*	*202*
,,	*215—218*	*59*	*48*
,,	*218—221*	*18*	*21*
,,	*221—224*	*2*	*5*

SOLUTION. Calculation of the Mean and the Standard Deviation of the expenditure.

Place A

Range of expenditure in Rs. per month	M. values	No. of families	Deviation from the assumed average (213.5)	dx/i where —i represents the common factor	Square of deviation	Frequency × deviation	Frequency × square of deviation
	(m)	(f)	(dx)	(dx)	dx^2	fdx	fdx^2
203—206	204.5	28	—9	—3	9	—84	252
206—209	207.5	292	—6	—2	4	—584	1168
209—212	210.5	389	—3	—1	1	—389	389
212—215	213.5	212	0	0	0	0	0
215—218	216.5	59	+3	+1	1	+ 59	59
218—221	219.5	18	+6	+2	4	+ 36	72
221—224	222.5	2	+9	+3	9	+ 6	18
		$n=1000$				$\Sigma fdx=-956$	$\Sigma fdx^2=1958$

Place B

203—206	204.5	39	—9	—3	9	—117	351
206—209	207.5	284	—6	—2	4	—568	1136
209—212	210.5	401	—3	—1	1	—401	401
212—215	213.5	202	0	0	0	0	0
215—218	216.5	48	+3	+1	1	+ 48	48
218—221	219.5	21	+6	+2	4	+ 42	84
221—224	222.5	5	+9	+3	9	+ 15	45
		$n=1000$				$\Sigma fdx=-981$	$\Sigma fdx^2=2065$

Place A

$$\text{Mean expenditure} = x + \frac{\Sigma fdx}{n} \times i = 213.5 + \frac{-956}{1000} \times 3 = 210.63 \text{ rupees.}$$

$$\text{Standard Deviation of the expenditure} = \sqrt{\frac{\Sigma f\,dx^2}{n} - \left(\frac{fdx}{n}\right)^2} \times i$$

$$= \sqrt{\frac{1958}{1000} - \left(\frac{-956}{1000}\right)^2} \times 3$$

$$= \sqrt{1.044} \times 3$$

$$= 1.02 \times 3 = 3.06 \text{ rupees}$$

Place B

$$\text{Mean expenditure} = x + \frac{\Sigma fdx}{n} \times i = 213.5 + \frac{-981}{1000} \times 3 = 210.56$$

$$\text{Standard Deviation of the expenditure} = \sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i$$

$$= \sqrt{\frac{2065}{1000} - \left[\frac{-981}{1000}\right]^2} \times 3$$

$$= \sqrt{1.103} \times 3$$

$$= 1.049 \times 3 = 3.15 \text{ rupees.}$$

PROBLEM. 2.26. *Find the mean yield of paddy and the Standard Deviations for the distribution of the results of 3061 crop cutting experiments shown in the following table :—*

Yield of paddy per acre in Kgs.	*No. of experiments*
0— 400	*236*
401— 800	*481*
801—1200	*604*
1201—1600	*576*
1601—2000	*419*
2001—2400	*333*
2401—2800	*217*
2801—3200	*87*
3201—3600	*64*
3601—4000	*23*
4001—4400	*14*
4401—4800	*6*
4801—5200	*1*
Total	*3061*

SOLUTION.

Calculation of the mean yield of paddy and the standard deviation as a result of 3061 crop-cutting experiments

Yield of paddy per acre in kgs. (*m*)	M. Values (*mv*)	No. of experiments (*f*)	Deviation as. av. *dx*	from the (1800.5) dx/i, where $i=400$	Deviation × Frequency (*fdx*)	Square of Deviation (dx^2)	Square of Deviation Frequency (fdx^2)
0— 400	200.5	236	—1600	—4	— 944	16	3776
401— 800	600.5	481	—1200	—3	—1443	9	4329
801 —1200	1000.5	604	— 800	—2	—1208	4	2416
1201—1600	1400.5	576	— 400	—1	— 576	1	576
1601 —2000	1800.5	419	0	0	0	0	0
2001—2400	2200.5	333	+ 400	+1	+ 333	1	333
2401—2800	2600.5	217	+ 800	+2	+ 434	4	868
2801—3200	3000.5	87	+1200	+3	+ 261	9	783
3201—3600	3400.5	64	+1600	+4	+ 256	16	1024
3601—4000	3800.5	23	+2000	+5	+ 115	25	575
4001—4400	4200.5	14	+2400	+6	+ 84	36	504
4401—4800	4600.5	6	+2800	+7	+ 42	49	294
4801—5200	5000.5	1	+3200	+8	+ 8	64	64
		$n=3061$			$\Sigma fdx=—2638$		$\Sigma fd^2x=15542$

Mean yield of paddy$=x+\frac{\Sigma fdx}{n}\times i$

$$=1800{\cdot}5+\frac{-2638}{3061}\times 400$$

$$=1455{\cdot}8 \text{ Kgs. per acre}$$

Standard Deviation of the yield of paddy for 3061 crop cutting experiment $=\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\times i$

$$=\sqrt{\frac{15542}{3061}-\left(\frac{-2638}{3061}\right)^2}\times 400$$

$$=\sqrt{4{\cdot}34}\times 400=2{\cdot}08\times 400=832 \text{ Approx.}$$

PROBLEM 2.27. *Find out of the mean and Standard Deviation of the series by the method of summation.*

Marks	*No. of Students*
0—10	*10*
10—20	*15*
20—30	*25*
30—40	*25*
40—50	*10*
50—60	*10*
60—70	*5*
	100

SOLUTION. Calculation of mean and standard deviation by the method of summation.

Mark	No. of Students	1st Cumulation	2nd Cumulation
0—10	10	10	10
10—20	15	25	35
20—30	25	50	85
30—40	25	75	160
40—50	10	85	245
50—60	10	95	340
60—70	5	100	440
	100	440	1315

Let M be the middle point of the last class intervals. In this case it is 65. Let i be the magnitude of class-intervals. In this case it is 10.

Let $\frac{\text{sum of 1st Cumulation}}{\Sigma f}=\frac{440}{100}=F_1=4.4$

and $\frac{\text{sum of 2nd Cumulation}}{\Sigma f}=\frac{1315}{100}=F_2=13.15$

Mean $=M-i\,(F_1-1)$

$=65-10\,(4{\cdot}4-1)=65-10\,(3{\cdot}4)=31$

Standard Deviation $=i\sqrt{F_2-F_1-F_1{}^2}=10\sqrt{26{\cdot}3-4.4-19{\cdot}36}$

$=10\sqrt{2{\cdot}54}$

$=10\times 1{\cdot}596=15{\cdot}96$ marks.

PROBLEM 2.28. *Find out the mean and Standard Deviation of the following data :—*

Age Under	*Number of persons dying*
10	*15*
20	*30*
30	*53*
40	*75*
50	*100*
60	*110*
70	*115*
80	*125*

SOLUTION. Calculation of the Mean and Standard Deviation.

Age (*m*)	No. of persons Dying (*f*)	Step Deviations from as. av. (35) (d')	(fd')	(fd'^2)
0—10	15	—3	—45	135
10—20	15	—2	—30	60
20—30	23	—1	—23	23
30—40	22	0	0	0
40—50	25	+1	+25	25
50—60	10	+2	+20	40
60—70	5	+3	+15	45
70—80	10	+4	+40	160
	$\Sigma f=125$		$\Sigma fd'=+2$	$\Sigma fd'^2=498$

$$\text{Mean or a} = x+\left(\frac{\Sigma fd'}{n}\times i\right)$$

$$=35+\left(\frac{2}{125}\times 10\right)$$

$$=35+\cdot 16 \qquad =35\cdot 16 \text{ years}$$

$$\text{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma fd'^2}{n}-\left(\frac{fd'}{n}\right)^2}\times i$$

$$=\sqrt{\frac{488}{125}-\left(\frac{2}{125}\right)^2}\times 10$$

$$=1\cdot 97\times 10.$$

$$=19.7 \text{ years.}$$

PROBLEM 2.29. *Find out the coefficient of variation of the following series—*

Income		*Number of persons*
More than	*1000*	*0*
,, ,,	*900*	*50*
,, ,,	*800*	*110*
,, ,,	*700*	*200*
,, ,,	*600*	*400*
,, ,,	*500*	*600*
,, ,,	*400*	*700*
,, ,,	*300*	*850*
,, ,,	*200*	*900*
,, ,,	*100*	*1000*

SOLUTION. Calculation of the Co-efficient of Variation.

Income (m)	No. of persons (f)	Step Deviation from as. av. (550) (d') $(i=100)$	fd'	fd'^2
900—1000	50	+4	+200	800
800— 900	60	+3	+180	540
700— 800	90	+2	+180	360
600— 700	200	+1	+200	200
500— 600	200	0	0	0
400— 500	100	—1	—100	100
300— 400	150	—2	—300	600
200— 300	50	—3	—150	450
100— 200	100	—4	—400	1600
	$n=1{,}000$		$\Sigma fd'=-190$	$\Sigma fd'^2=4650$

Mean or $a=x+\left(\frac{\Sigma fd'}{n}\times i\right)=550+\left(\frac{-190}{1000}\times 100\right)$

$=531$ rupees

Standard Deviation or σ $=\sqrt{\frac{\Sigma fd'^2}{n}-\left(\frac{\Sigma fd'^2}{n}\right)^2}\times i$

$=\sqrt{\frac{4650}{1000}-\left(\frac{-190}{1000}\right)^2}\times 100$

$=\sqrt{4.6139}\times 100$

$=2.147\times 100=214.7$ rupees

Co-efficient of variation $=\frac{\sigma}{a}\times 100=\frac{214.7}{531}\times 100=40.45$

PROBLEM 2.30. *Calculate the standard deviation of the following two serie Which shows greater deviation ?*

Series A	*Series B*
192	*83*
288	*87*
236	*93*
229	*109*
184	*124*
260	*126*
348	*126*
291	*101*
330	*102*
243	*108*

SOLUTION Calculation of the Standard Deviations of series A and B.

Series A			Series B		
Size of the item (m)	Deviation from arith. av. (260.1) (d)	Square of Deviation (d^2)	Size of the item (m)	Deviation from arith. av. (105.9) (d)	Square of Deviation (d^2)
192	−68.1	4637.61	83	−22.9	524.41
288	+27.9	778.41	87	−18.9	357.21
236	−24.1	580.81	93	−12.9	166.41
229	−31.1	967.21	109	+ 3.1	9.61
184	−76.1	5791.21	124	+18.1	327.61
260	− .1	.01	126	+20.1	404.01
348	+87.9	7726.41	126	+20.1	404.01
291	+30.9	954.81	101	− 4.9	24.01
330	+69.9	4886.01	102	− 3.9	15.21
243	−17.1	292.41	108	+ 2.1	4.41
$m=2601$ $n=10$		$\Sigma d^2=$ 26614.90	$\Sigma m+1059$		$\Sigma d^2=$ 2236.90

eries A —

Arithmetic average or $a=\frac{\Sigma m}{n}=\frac{2601}{10}=260.1$

Standard Deviation or $\sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{26614.90}{10}}=51.6$

Co-efficient of Standard Deviation$=\frac{\sigma}{a}=\frac{51.6}{260.1}=.198$

ries B —

Arithmetic average or $a=\frac{\Sigma m}{n}=\frac{1059}{10}=105.9$

Standard Deviation or $\sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{2236.90}{10}}=14.95$

Co-efficient of Standard Deviation$=\frac{\sigma}{a}=\frac{14.95}{105.9}=.141$

On comparison of standard co-efficients of deviation, it is obvious that series hows greater deviation.

PROBLEM 2.31 *The index numbers of prices of cotton and coal shares in 1972 e as under —*

Month	*Index number of prices of cotton shares*	*Index number of prices of coal shares*
January	*188*	*131*
February	*178*	*130*
March	*173*	*130*
April	*164*	*129*

May	*172*	*129*
June	*183*	*120*
July	*184*	*127*
August	*185*	*127*
September	*211*	*130*
October	*217*	*137*
November	*232*	*140*
December	*240*	*142*

Which of the two shares do you consider more variable in price ?

SOLUTION. Calculation of Standard Deviation.

Month	Cotton Shares			Coal Shares		
	Index number of prices (m)	Deviation from mean (194) (d)	Deviation Squared up (d^2)	Index number of prices (m)	Deviation from mean (131) (d)	Devi tion Square up (d^2)
January	188	-6	36	131	0	
February	178	-16	256	130	-1	
March	173	-21	441	130	-1	
April	164	-30	900	129	-2	
May	172	-22	484	129	-2	
June	183	-11	121	120	-11	
July	184	-10	100	127	-4	
August	185	-9	81	127	-4	
September	211	$+17$	289	130	-1	
October	217	$+23$	529	137	$+6$	
November	232	$+38$	1444	140	$+9$	
December	240	$+46$	2116	142	$+11$	
$n=12$	$\Sigma m=2327$		$\Sigma d^2=$ 6797	$\Sigma m=1572$		$\Sigma d^2=$

Cotton Shares

$$\text{Arithmetic average or } a=\frac{\Sigma m}{n}=\frac{2327}{12}=194 \text{ approx.}$$

$$\text{Standard Deviation or } \sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{6797}{12}}=23.8$$

$$\text{Co-efficient of Variation}=\frac{\sigma}{a}\times 100=\frac{23.8}{194}\times 100=12.27$$

Coal Shares

$$\text{Arithmetic average or } a=\frac{\Sigma m}{n}=\frac{1572}{12}=131$$

$$\text{Standard Deviation or } \sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{402}{12}}=\sqrt{33.5}=5{\cdot}79$$

$$\text{Co-efficient of Variation}=\frac{\sigma}{a}\times 100=\frac{5.79\times 100}{131}=4.42$$

A comparison of the co-efficients of variation of cotton and coal shares reveals that cotton shares are more variable in prices than the coal shares.

PROBLEM 2.32 *Following are the marks obtained by two students A and B in 10 sets of examinations :*

Sets	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Marks obtained by A	*44*	*80*	*76*	*48*	*52*	*72*	*68*	*56*	*60*	*54*
Marks obtained by B	*48*	*75*	*54*	*60*	*63*	*69*	*72*	*51*	*57*	*66*

If the consistency of performance is the criterion for awarding a prize, who should get the prize ?

SOLUTION. Calculation of co-efficients of variation

Sets	A			B		
	Marks (m)	Deviations from 60 (dx)	Square of deviations (dx^2)	Marks (m)	Diviations from 60 (dx)	Square of deviation (dx^2)
1	44	-16	256	48	-12	144
2	80	$+20$	400	75	$+15$	225
3	76	$+16$	256	54	-6	36
4	48	-12	144	60	0	0
5	52	-8	64	63	$+3$	9
6	72	$+12$	144	69	$+9$	81
7	68	$+8$	64	72	$+12$	144
8	56	-4	16	51	-9	81
9	60	0	0	57	-3	9
10	54	-6	36	66	$+6$	36
	$n=10$	$\Sigma dx=10$	$\Sigma dx^2=1380$	$n=10$	$\Sigma dx^1=15$	$\Sigma dx^2=765$

A. Arithmetic Average a $=60+\frac{10}{10}=61$ marks.

Standard Deviation or $\sigma=\sqrt{\frac{1380}{10}-\left(\frac{10}{10}\right)^2}$

$=\sqrt{138-1}=\sqrt{137}=11{\cdot}7$ approx.

Co-efficient of Variation $=\frac{\sigma}{a}\times 100=\frac{11\cdot 7}{61}\times 100=19{\cdot}18$

B. Arithmetic Average or a $=60+\frac{15}{10}=61{\cdot}5$ marks.

Standard Deviation or $\sigma=\sqrt{\frac{765}{10}-\left(\frac{15}{10}\right)^2}$

$=\sqrt{76{\cdot}5-2{\cdot}25}=\sqrt{74{\cdot}25}=8.62$

Co-efficient of Variation $=\frac{8\cdot 62}{61\cdot 5}\times 100=14{\cdot}02$

B is more consistent in securing the marks, hence the prize should be given to B.

PROBLEM 2.33 *The following table gives the area under cultivation, production, net export, and price of wheat in the U. S. A. :—*

Year	*Area under cultivation (millions of Acres)*	*Production (millions of Bushel)*	*Net export (millions of Bushel)*	*Price per Bushel in cents*
1890	*36.7*	*449*	*109*	*84*
1895	*38.9*	*542*	*130*	*51*
1900	*49.2*	*599*	*221*	*62*
1905	*46.3*	*706*	*101*	*75*
1910	*45.8*	*626*	*70*	*91*
1915	*60.3*	*1008*	*240*	*96*
1920	*62.3*	*843*	*313*	*183*
1925	*52.4*	*669*	*93*	*144*
1930	*62.6*	*886*	*112*	*67*
1935	*51.3*	*626*	*0*	*83*

Compare the co-efficient of variation of area under cultivation and net export of wheat for the whole period.

SOLUTION. Calculation of the Standard Deviation of Area under cultivation and Net Exports.

Year	Area under cultivation			Net Exports		
	Area in millions Acres (m)	Deviation from the average (50.6) (d)	Square of Deviation (d^2)	Exports in million of Bushel (m)	Deviation from Average (d)	Square of deviation (d^2)
1890	36.7	—13.9	193.21	109	—30	900
1895	38.9	—11.7	136.89	130	— 9	81
1900	49.2	— 1.4	1.96	221	+82	6724
1905	46.3	— 4.3	18.49	101	—38	1444
1910	45.8	— 4.8	23.04	70	—69	4761
1915	60.3	+ 9.7	94.09	240	+101	10201
1920	62.3	+11.7	136.89	313	+174	30276
1925	52.4	+ 1.8	3.24	93	— 46	2116
1930	62.6	+12.0	144.00	112	— 27	729
1935	51.3	+ .7	.49	0	—139	19321
$n=10$	$\Sigma m=505.8$		$\Sigma d^2+752.30$	$\Sigma m—1389$		$\Sigma d^2—76554$

Area

Average or $a = \frac{\Sigma m}{n} = \frac{505.8}{10}$ =50.6 million acres approx.

Standard Deviation or $\sigma = \sqrt{\frac{\Sigma d^2}{n}} = \sqrt{\frac{752.30}{10}} = \sqrt{75.23} = 8.674$ million acres

$$\text{Co-efficient of Variation} = \frac{\sigma}{a} \times 100 = \frac{8.674}{50.6} \times 100$$

$$= 17.14$$

Exports

$$\text{Average or } a = \frac{\Sigma m}{n} = \frac{1389}{10} = 139 \text{ million of bushels}$$

$$\text{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma d^2}{n}} = \sqrt{\frac{76553}{10}} = \sqrt{7655.3}$$

$$= 87.49 \text{ million of bushels.}$$

$$\text{Co-efficient of Variation} = \frac{\sigma}{a} \times 100 = \frac{87.49}{139} \times 100$$

$$= 62.94$$

PROBLEM 2.34 *In any two series, where d_1 and d_2 represent the deviation from a trial average, 100.*

$n_1 = 150$	$\Sigma d_1 = 180$	$\Sigma d_1^2 = 245320$
$n_2 = 200$	$\Sigma d_2 = 250$	$\Sigma d_2^2 = 43850$
	$a = 100$	

Calculate the co-efficient of variation for the two series.

SOLUTION.

Series 1

$$\text{Arithmetic average or } a = x + \frac{\Sigma d_1}{n_1} = 100 \times \frac{180}{150} = 101.2$$

$$\text{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma d_1^2 - n(a-x)^2}{n}}$$

$$= \sqrt{\frac{245320 - 150(1{\cdot}2)^2}{150}}$$

$$\sqrt{1634{\cdot}03}$$

$$= 40{\cdot}4$$

$$\text{Co-efficient of Variation} = \frac{\sigma}{a} \times 100 = \frac{40{\cdot}4}{101{\cdot}2} \times 100$$

$$= 39{\cdot}92$$

Series 2

$$\text{Arithmetic average or } a = x + \frac{\Sigma d_2}{n_2} = 100 + \frac{250}{200}$$

$$= 101{\cdot}25$$

$$\text{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma d_2^2 - n\,(a-x)^2}{n}}$$

$$= \sqrt{\frac{43850 - 200(1{\cdot}25)^2}{200}} = \sqrt{217{\cdot}69}$$

$$= 14{\cdot}75$$

$$\text{Co-efficient of Variation} = \frac{\sigma}{a} \times 100 = \frac{14{\cdot}75}{101{\cdot}25} \times 100$$

$$= 14{\cdot}6$$

PROBLEM 2.35 *In any two samples, where the variates x_1 and x_2 are measured in the same units,*

$n_1=36$ (*summation*) $\Sigma x_1^2=49428$

$n_2=49$ (*summation*) $\Sigma x_2^2=71258$

Compute the values of the Standard Deviation of the two samples. What additional information is required to calculate the co-efficient of variation of the above two samples ? Indicate the uses of such a co-efficient.

SOLUTION.

$$\text{Standard Deviation of the first sample} = \sqrt{\frac{\Sigma x_1^2}{n_1}} = \sqrt{\frac{49428}{36}} = \sqrt{1373} = 37{\cdot}06$$

$$\text{Standard Deviation of the 2nd sample} = \sqrt{\frac{\Sigma x_2^2}{n_2}} = \sqrt{\frac{71258}{49}} = \sqrt{1454{\cdot}25} = 38{\cdot}13$$

The values of a_1 and a_2 (where a represents the arithmetic average) will also be required if the calculation of the co-efficient of variation of the above two samples is also desired.

PROBLEM 2.36. *Compile a table, showing the frequencies with which words of different numbers of letters occur in the extract reproduced below (omitting punctuation marks) treating as the variable the number of letters in each word, and obtain the mean, median, and the co-efficient of variation of the distribution :—*

"Success in the examination confers no absolute right to appointment, unless Government is satisfied, after such enquiry as may be considered necessary, that the candidate is suitable in all respects for appointment to the public service."

SOLUTION. Calculation of the mean, median and the co-efficient of variation.

Size of item (m)	Frequency (f)	Deviation from the assumed average (6) dx	Total Deviations fdx	Cumulative Frequency cf	dx^2	fdx^2
2	9	—4	—36	9	16	144
3	6	—3	—18	15	9	54
4	2	—2	—4	17	4	8
5	2	—1	—2	19	2	4
6	2	0	0	21	0	0
7	4	+1	+4	25	1	4
8	3	+2	+6	28	4	12
9	3	+3	+9	31	9	27
10	2	+4	+8	33	16	32
11	3	+5	+15	36	25	75
	$n=36$		$\Sigma fdx=-18$			$\Sigma fdx^2=360$

Mean$=x+\frac{\Sigma fdx}{n}=6+\frac{-18}{36}=5.5$

Median = the size of the $\frac{n+1}{2}$ that is, 18.5th item

$=5$

Standard Deviation$=\sqrt{\frac{\Sigma fdx^2-n(a-x)^2}{n}}$

$=\sqrt{\frac{360-36\,(.5)^2}{36}}$

$=\sqrt{\frac{360-36\,(.25)}{36}}$

$=\sqrt{9.75}$

$=3.12$

Co-efficient of Variation $=100\times\frac{\text{Standard Deviation}}{\text{Arithmetic Average}}$

$=100\times\frac{3.12}{5.5}$

$=56.7$

PROBLEM 2.37. *What are measures of dispersion of a distribution ? Why is the standard deviation most commonly used as a measure of dispersion in statistics .*

Goals scored by two teams A and B in a football season were as follows :

Number of goals scored in a match	*Number of Matches*	
	A	*B*
0	*27*	*17*
1	*9*	*9*
2	*8*	*6*
3	*5*	*5*
4	*4*	*3*

By calculating the co-efficient of variation in each case find which team may be considered more consistent.

SOLUTION. Calculation of co-efficient of variation.

No. of goals (m)	Deviation from as. av. (2). (dx)	Square of Deviation (dx^2)	A			B		
			(f)	Frequency × Deviation (fdx)	Frequency × square of Deviation (fdx^2)	(f)	Frequency × Deviation (fdx)	Frequency × Square of Deviation (fdx^2)
0	−2	4	27	−54	108	17	−34	68
1	−1	1	9	−9	9	9	−9	9
2	0	0	8	0	0	6	0	0
3	+1	1	5	+5	5	5	+5	5
4	+2	4	4	+8	16	3	+6	12
			$n=53$	$\Sigma fdx=-50$	$\Sigma fdx^2=138$	$n=40$	$\Sigma fdx=-32$	$\Sigma fdx^2=94$

Team A

Average goals scored or $a = x + \frac{\Sigma fdx}{n} = 2 + \frac{-50}{53} = 1.06$

Standard Deviation or $\sigma = \sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2}$

$$= \sqrt{\frac{138}{53} - \left(\frac{-50}{53}\right)^2}$$

$$= \sqrt{1.71}$$

$$= 1.3$$

Co-efficient of variation $= \frac{\sigma}{a} \times 100$

$$= \frac{1.3}{1.06} \times 100 = 122.6$$

Team B

Average goals scored or $a = x + \frac{\Sigma fdx}{n} = 2 + \frac{-32}{40} = 1.2$

Standard Deviation or $\sigma = \sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2}$

$$= \sqrt{\frac{94}{40} - \left(\frac{-32}{40}\right)^2}$$

$$= \sqrt{1.71}$$

$$= 1.3$$

Co-efficient of variation $= \frac{\sigma}{a} \times 100$

$$= \frac{1.3}{1.2} \times 100 = 108.3$$

It is now clear that team B is more consistent than team A.

PROBLEM 2.38. *The following table gives the marks obtained by 10 students :—*

Digits (Division of Class— interval)

Marks	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Total*
0— 9	*2*		*4*		*3*	*1*			*1*	*1*	*12*
10—19	*5*	*3*		*4*		*2*			*1*		*15*
20—29		*1*	*7*	*8*	*10*	*5*		*4*	*3*	*2*	*40*
30—39			*3*	*5*	*10*	*2*		*1*		*1*	*22*
40—49			*4*	*3*	*2*		*2*				*11*
											100

Thus, 4 marks are obtained by 3 students, 13 marks by 4 students, 35 marks by 2 students and so on.

Calculate the mean marks and standard deviation of marks :

(i) By using the totals only.

(ii) By using the whole data.

SOLUTION. Calculation of Mean and Standard Deviation

Digits (Division of Class —intervals)

Marks	0	1	2	3	4	5	6	7	8	9	Total (*f*)	Step Dev. from (25) (*d*)	*fd'*	*fd'*²
0—9	0 2 0		8 4 16		12 3 48	5 1 25			8 1 64	9 1 81	42 12 234	—2	—24	48
10—19	50 5 500	33 3 363		52 4 676		30 2 450			18 1 324		183 15 2313	—1	—15	15
20—29		21 1 441	154 7 3388	184 8 4232	240 10 5760	125 5 3125		108 4 2916	84 3 2352	58 2 1682	974 40 23896	0	0	0
30—39			96 3 3072	165 5 5445	340 10 11560	70 2 2450		37 1 1369		39 1 1521	747 22 25417	+1	+22	22
40—49			168 4 7056	129 3 5547	88 2 3872		92 2 4232				477 11 20707	+2	+22	44
								Total			2423 100 72567		+5	129

(*a*) *Using the totals only* :

$$\text{Mean or } a = x + \left(\frac{\Sigma fd'}{n} \times i\right)$$

$$= 25 + \left(\frac{5}{100} + 10\right) = 25.5 \text{ marks}$$

$$\text{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma fd'^2}{n} - \left(\frac{\Sigma fd'}{n}\right)^2} \times i$$

$$= \sqrt{\frac{129}{100} - \left(\frac{5}{100}\right)^2} \times 10 = 11.3 \text{ marks.}$$

(*b*) *Using the whole data* :

Mean $= \frac{\Sigma mf}{n}$ where m stands for the value of an item and f for frequency.

Figures in the upper left corner of each square in the table indicate 'mf'.

$$= \frac{2423}{100} = 24.23 \text{ marks}$$

$$\text{Standard Deviation} = \sqrt{\frac{\Sigma m^2f - (mf)^2/n}{n}}$$

Figures in the lower right corner of each square in the table indicate '$m^2 f$'

$$= \sqrt{\frac{72567 - (2423)^2/100}{100}}$$

$$= \sqrt{\frac{72567 - 58709.29}{100}}$$

$$= \sqrt{138.5771}$$

$$= 11.77 \text{ marks.}$$

PROBLEM 2.39. *A collar manufacturer is considering the production of a new style of collar to attract young men. The following statistics of neck circumferences are available based upon measurements of a typical group of college students :—*

Mid-value (centimetres)	*No. of Students*
32	*4*
33	*19*
34	*30*
35	*63*
36	*66*
37	*29*
38	*18*
39	*1*
40	*1*

Compute the Standard Deviation and use the criterion ($X \pm 3$ Standard Deviation) to determine the largest and smallest size of collars he should make in order to meet the needs of practically all his customers, bearing in mind that collars are worn, on average, 1 centimetre larger than neck size.

SOLUTION. Computation of the Mean and the Standard Deviation of the neck-circumferences of 231 students.

Neck circumference in centimetres (m)	No. of students (f)	Deviation from the assumed average (35) (dx)	Deviation × Frequency (fdx)	Square of deviation (dx^2)	Square of deviation × Frequency (fdx^2)
32	4	—3	—12	9	36
33	19	—2	—38	4	76
34	30	—1	—30	1	30
35	63	—0	— 0	0	0
36	66	—1	—66	1	66
37	29	+2	+58	4	116
38	18	+3	+54	9	162
39	1	+4	+ 5	16	16
40	1	+5	+ 4	25	25
	$n=231$		$\Sigma fdx=+107$		$\Sigma fdx^2=527$

Mean of the neck circumference $=x+\frac{\Sigma fdx}{n}=35+\frac{107}{231}$

$=35.46$ centimetres

Standard Deviation of the neck-circumferences

$$=\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}$$

$$=\sqrt{\frac{527}{231}-\left(\frac{107}{231}\right)^2}$$

$$=\sqrt{2.07}$$

$=1.439$ centimetres

Using the criterion $\mathbf{X}+3$ Standard Deviation,

(1) The smallest size of the collars $=\mathbf{X}-3\sigma+1$ centimetres

$=35.46-3\,(1.439)+1.00$ centimetres

$=32.14$ centimetres.

(2) The largest size of the collars $=\mathbf{X}+3\sigma+1$ centimetres

$=35.46+3\,(1.439)+1.00$ centimetres

$=40.78$ centimetres

PROBLEM 2.40. *Calculate the arithmetic average and the standard deviation of following figuresand state the percentage of case which lies outside the mean at distance* $a\pm\sigma$, $a\pm2\sigma$, $a\pm3\sigma$, *where* σ *stands for the standard deviation.*

148, 145, 141, 116, 96, 91, 87, 89, 91, 91, 102, 95, 108, 120, 139.

SOLUTION. Calculation of Standard Deviation

Size of item	Deviation from assumed average (100) (dx)	Square of Deviatio (d^2x)
148	+48	2304
145	+45	2025
141	+41	1681
116	+16	256
96	— 4	16
91	— 9	81
87	—13	169
89	—11	121
91	— 9	81
91	— 9	81
102	+ 2	4
95	— 5	25
108	+ 8	64
120	+20	400
139	+39	1521
Total $n=15$	$\Sigma dx=+159$	$\Sigma d^2x=8829$

$$\text{Mean} = x + \frac{\Sigma dx}{n} = 100 + \frac{159}{15} = 110.6$$

$$\text{Standard Deviation, or } \sigma = \sqrt{\frac{\Sigma d^2x}{n} - \left(\frac{\Sigma dx}{n}\right)^2}$$

$$= \sqrt{\frac{8829}{15} - \left(\frac{159}{15}\right)^2}$$

$$=21.82 \text{ nearly.}$$

(1) Cases lying outside the mean at distance $\alpha \pm \sigma$, that is outside the (89—132) group.

$=5.$

Therefore, the percentage of cases $= \frac{5}{15} \times 100 = 33.3$

(2) Cases lying outside the mean at distance $\alpha \pm 2\sigma = 0$
Therefore, the percentages of cases $=0$.

(3) Cases lying outside the mean at distance $\alpha \pm 3\sigma = 0$
Therefore, the percentages of cases$=0$.

PROBLEM 2.41. *For a certain group of 'saree' weaver of Banaras, the med and quartile earnings per week are Rs. 144.3, Rs. 143.0, and Rs. 145.9 respectiv The earnings for the group range between Rs.140 and Rs 150. Ten per cent of t group earn under Rs142 per week, 13 per cent earn Rs147 and over and 6 per ce Rs148 and over. Put these data into the form of the frequency distribution and obt an estimate of the mean wage and the standard deviation.*

SOLUTION On the supposition that the group consists of 100 persons t following type of frequency distribution can be obtained, and its mean and stand deviation can be calculated.

Earnings	Mid-points (m)	No. of weavers (f)	(mf)	Deviations actual av. (d)144.5	d^2	fd^2
Rs.						
140 —142	141.00	10	1410.00	—3.50	12.25	122.5
142 —143	142.50	15	2137.50	—2.00	4.00	60.0
143 —144.3	143.65	25	3591.25	— .85	.72	18.0
144.3—145.9	145.10	25	3627.50	+ .60	.36	9.0
145.9—147	146.45	12	1757.40	+1.95	3.80	45.6
147 —148	147.50	7	1032.50	+3.00	9.00	63.0
148 —150	149.00	6	894.00	+4.50	20.25	121.5
		n—100	Σmf 14450.15			Σfd^2 439.6

$$\text{Mean} = \frac{\Sigma mf}{\Sigma f} = \frac{14450.15}{100} = 144.5 \text{ rupees.}$$

$$\text{Standard Deviation} = \sqrt{\frac{\Sigma fd^2}{n}} = \sqrt{\frac{439.6}{100}} = 2.097 \text{ rupees.}$$

PROBLEM 2.42. *The following table relates to the profit and losses of 100 firms. Calculate the average profits and the standard deviation of profits.*

Profits Rs.	*Number Of Firms*
5000 to 6000	*8*
4000 to 5000	*12*
3000 to 4000	*30*
2000 to 3000	*10*
1000 to 2000	*5*
0 to 1000	*5*
—1000 to 0	*6*
—2000 to —1000	*8*
—3000 to —2000	*9*
—4000 to —3000	*7*

SOLUTION. Calculation of Mean profits and Standard Deviation of profits.

Profits (m)	No. of Firms (f)	Step. Deviation from as. av. (1500) (d′) (i=1000)	fd'	fd'^2
5000 to 6000	8	+4	+32	128
4000 to 5000	12	+3	+36	108
3000 to 4000	30	+2	+60	120
2000 to 3000	10	+1	+10	10
1000 to 2000	5	0	0	0
0 to 1000	5	—1	—5	5
—1000 to 0	6	—2	—12	24
—2000 to —1000	8	—3	—24	72
—3000 to —2000	9	—4	—36	144
—2000 to —1000	7	—5	—35	175
	n=100		$\Sigma fd' = +26$	$\Sigma fd'^2 = 786$

$$\text{Mean or } a = x + \left(\frac{\Sigma fd'}{n} \times i\right)$$
$$= 1500 + \left(\frac{26}{100} \times 1000\right)$$
$$= 1500 + 260.$$
$$= 1760 \text{ rupees.}$$

$$\text{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma fd'^2}{n} - \left(\frac{\Sigma fd'}{n}\right)^2} \times i$$
$$= \sqrt{\frac{786}{100} - \left(\frac{26}{100}\right)^2} \times 1000$$
$$= \sqrt{7.79 \times 1000}$$
$$= 2.791 \times 1000$$
$$= 2791 \text{ rupees.}$$

PROBLEM 2.43 *Explain clearly the ideas implied in using arbitrary work origin and scale for the calculation of the arithmetic mean and Standard Deviat of a frequency distribution.*

The values of the arithmetic mean and Standard Deviation of the followi frequency distribution of a continuous variable derived from analysis in the abo manner are 135.3 £ and 96 £ respectively.

Deviation	—4	—3	—2	— 1	0	+1	+2	+3	*Total*
frequency	2	5	8	18	22	13	8	4	80

Determine the actual class-intervals.

SOLUTION

Step Deviation (d)	Frequency (f)	(fd)	(fd^2)
—4	2	— 8	32
—3	5	—15	45
—2	8	—16	32
—1	18	—18	18
0	22	0	0
+1	13	+13	13
+2	8	+16	32
+3	4	+12	36
	$n=80$	$\Sigma fd = -16$	$\Sigma fd^2 = 208$

Standard Deviation as given is 96

$$\text{or } \sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} \times i = 96$$
$$= \sqrt{\frac{208}{80} - \left(\frac{-16}{80}\right)^2} \times i = 96$$
$$= 1.6 \times i = 96$$
$$\therefore i = \frac{96}{1.6} = 60$$

The magnitude of the class intervals is thus 60

Arithmetic Average as given is 135.3

$$\text{or } a+\left(\frac{\Sigma fd}{n} \times i\right)=135.3$$

$$=a+\left(\frac{-16}{80} \times 60\right)=135.3$$

$$a-12=135.3$$

$$a=147.3$$

Thus the assumed average is 147.3.

Since the magnitude of the class intervals is 60 and the assumed average is 147.3, therefore the class interval of the assumed average would be (147.3±30) 117.3—177.3. Other class intervals would be as follows :—

Class Interval	Frequency	Class Interval	Frequency
−122.7 to — 62.7	2	117.3 to 177.3	22
— 62.7 to — 2.7	5	177.3 to 237.3	13
— 2.7 to 57.3	8	237.3 to 297.3	8
57.3 to 117.3	18	297.3 to 357.3	4

PROBLEM 2.44. *Mean and standard deviation of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the mean and standard deviation of all the 250 items taken together.*

SOLUTION :

Combined mean $(a_{12})=\frac{a_1 n_1+a_2 n_2}{n_1+n_2}$ where a_1 and a_2 stand for the mean of two series and n_1 and n_2 for the number of items in the two series. Substituting the values.

$$a_{12}=\frac{50\times 100+40\times 150}{100+150}=\frac{5000+6000}{250}=\frac{11,000}{250}=44$$

$$\text{Combined standard deviation } (\sigma_{12})=\sqrt{\frac{n_1\,(d_1{}^2+\sigma_1{}^2)+n_2\,(d_2{}^2+\sigma_2{}^2)}{n_1+n_2}}$$

Where σ_1 and σ_2 stand for the standard deviations of two series and $d_1=a_{12}-a_1$ and $d_2=a_{12}-a_2$ respectively. Substituting the values.

$$\sigma_{12}=\sqrt{\frac{100\,(36+25)+150\,(16+36)}{100+150}}$$

$$=\sqrt{\frac{100\times 61+150\times 52}{250}}$$

$$=\sqrt{\frac{6100+7800}{250}}$$

$$=7.457$$

PROBLEM 2.45. *An analysis of the monthly wages paid to workers in two firms A and B, belonging to the same industry, gives the following results :—*

	Firm A	*Firm B*
Number of wage-earners	*586*	*648*
Average monthly-wage	*Rs. 252.5*	*Rs. 247.5*
Variance of the distribution of wage	*100*	*121*

(*a*) *Which firm, A or B, pays out the larger amount as monthly wages ?*

(*b*) *In which firm, A or B, is there greater variability in individual wages ?*

(*c*) *What are the measures of* (*i*) *average monthly wage, and* (*ii*) *the variability in individual wages, of all the workers in the two firms, A and B, taken together.*

SOLUTION

Firm A

The average monthly wage per worker is Rs. 252.5
and the total no. of workers is 586
The total monthly wages paid to workers is Rs. 1,47,965

Firm B

The average monthly wage per worker is Rs. 247.5
The total no. of workers is 648
The total monthly wages paid to workers is Rs. 1,60,380

On comparison, it is clear that the Firm B pays out a larger amount as monthly wages.

Firm A

Variance of the distribution of wages is 100,
that is, the Standard Deviation $=\sqrt{100}=10$
and the arithmatic average $=252.5$

$\therefore$ Co-efficient of Standard Deviation $=\dfrac{\sigma}{a}$

of the distribution of wages

$$=\frac{10}{252.5}$$
$$=.0397$$

Firm B

Variance of the distribution of wages 121,
that is, the Standard Deviation $=\sqrt{121}=11$
the arithmetic average $=247.5$

$\therefore$ the Co-efficient of Standard Deviation $=\dfrac{\sigma}{a}$

of the distribution of wages

$$=\frac{11}{247.5}$$
$$=\cdot 044$$

On comparison, it is obvious that there is greater variability in individual wages in Firm B.

The average monthly wage of all the workers in both the factories A and B would be obtained by the following formula.

$m_{12}=\dfrac{n_1m_1+n_2m_2}{n_1+n_2}$ where

n_1 and n_2 stand for the number of items in the two series and m_1 and m_2 for the mean of the two series respectively and m_{12} for the combined mean of the two series. Substituting the values

$$\text{the combined average} = \frac{(586 \times 252.5)+(648 \times 247.5)}{586+648}$$

$$= \text{Rs. } \frac{3,08,345}{1234}$$

$$= \text{Rs. } 249.9.$$

The combined standard deviation of the two series would be given by the following formula :—

$$\sigma_{12} = \sqrt{\frac{n_1\sigma_1{}^2 + n_1\,(m_1 - m_{12})^2 + n_2\sigma_2{}^2 + n_2\,(m_2 - m_{12})^2}{n_1 + n_2}} \text{ where}$$

n_1 and n_2 and m_1, m_2 and m_{12} stand for the same things as explained above and σ_1, σ_2 and σ_{12} stand for the standard deviations of the 1st, 2nd and combined series respectively. Substituting the values,

$$\sigma_{12} = \sqrt{\frac{(586 \times 100)+586(252{,}5-249.9)^2+(648 \times 121)+648(247.5-249.9)^2}{586+648}}$$

$$= \sqrt{\frac{58600+(586 \times 6.76)+78408+(648 \times 5.76)}{1234}}$$

$$= \sqrt{\frac{144701.84}{1234}}$$

$$= \sqrt{117.26}$$

$$= \quad 10.8$$

PROBLEM 2.46. *From the data below, giving the averages and standard deviations of four sub-groups, calculate the average and the standard deviation of the whole group.*

Sub-group	*No. of men*	*Average wage* *Rs.*	*P.*	*S. D. wage* *Rs.*	*P.*
A	*50*	*61*	*0*	*8*	*0*
B	*100*	*70*	*0*	*9*	*0*
C	*120*	*80*	*5*	*10*	*0*
D	*30*	*83*	*0*	*11*	*0*
Total	*300*				

SOLUTION

Total wages of 50 men in sub-group A $= 50 \times 61 = 3050$ rupees
Total wages of 100 men in sub-group B $= 100 \times 70 = 7000$ rupees
Total wages of 120 men in sub-group C $= 120 \times 80.5 = 9660$ rupees
Total wages of 30 men in sub-group D $= 30 \times 83 = 2490$ rupees
Total wages of 300 men in sub-groups, A, B, C and D $= (3050+7000+9660+2490) = 22,200$ rupees

$$\text{Average of the whole group} = \frac{22,200}{300} = 74 \text{ rupees}$$

(ii) We know that $s^2=\sigma^2+d^2$

where, s^2 is the second moment about any arbitrary number and a^2 is the difference between the mean and this number.

Now mean of the whole group is 74.

$\therefore d_1=74-61=13 : d_2=74-70=4 : d_3=80.5-74=6.5$ and $d_4=83-74=9$

Now $N\sigma^2=N_1(\sigma_1^2+d_1^2)+N_2(\sigma_2^2+d_2^2)+N_3(c_3^2+d_3^2)+N_4(\sigma_4^2+d_4^2)$

or $300\,\sigma^2=50\,(64+169)+100\,(81+16)+120\,(100+42.25)+30\,(121+81)$

$=11650+9700+17070+6060=44{,}480$

or $\sigma^2 \quad =148.27$

$=12.1$

PROBLEM 2.47. *For a frequency distribution of marks in history of 200 candidates (grouped in intervals 0—5, 5—10, ...etc.), the mean and standard deviation were found to be 40 and 15. Later it was discovered that the score 43 was misread as 53 in obtaining the frequency distribution. Find the corrected mean and standard deviation corresponding to the corrected frequency distribution.*

SOLUTION. Let us reconstruct the original and wrong tables. Let the frequency of the ith group be fi then :

Class-intervals	Mid-values	Frequency	Frequency (correct)
0— 5	2.5	f_1	f_1
5—10	7.5	f_2	f_2
....	....	....	....
....	....	....	....
35—40	37.5	f_8	f_8
40—45	42.5	f_9	f_9+1
45—50	47.5	f_{10}	f_{10}
50—55	52.5	f_{11}	$f_{11}-1$
....	....	....	....
		200	200

The value of the mean when 43 was misread as 53 is given by

$$40=\frac{1}{200}\,(2.5f_1+7.5f_2+\ldots\ldots+37.5f_5+42.5f_9+47.5f_{10}+52.5f_{11}+\ldots\ldots)$$

Let the value of the corrected mean be x.

$$\text{Then } \overline{x}=\frac{1}{200}\,(2.5f_1+7.5f_2+\ldots+37.5f_8+42.5(f_9+1)+47.5f_{10}+52.5(f_{11}-1)+\ldots\ldots)$$

Let $2.5f_1+\ldots\ldots+37.5f_8+47.5f_{10}+57.5f_{12}+\ldots\ldots=s$

Then $40=\frac{1}{200}\,(s+42.5f+52.5_9f_{11})$ or $s+42.5f_9+52.5f_{11}=8000$

$$\text{and } \overline{x}=\frac{1}{200}\,[s+42.5(f_9+1)+52.5(f_{11}-1)]$$

$$=\frac{1}{200}\,[s+42.5f_9+52.5f_{11}+42.5-52.5]$$

$$=\frac{1}{200}\,[8000-10]=\frac{7990}{200}=39.95$$

The corrected mean corresponding to the corrected distribution is 39.95.

Calculation of the corrected Standard Deviation.

When 43 was misread as 53, the second moment about 40 which was thought o be equal to σ^2, is given by :

$$(15)^2=\frac{1}{200}[f_1(2.5-40)^2+f_2(7.5-40)^2+\ldots+f_9(42.5-40)^2+f_{10}$$
$$(47.5-40)^2+f_{11}(52.5-40)^2+\ldots\ldots]$$

Let $f_1\ 37.5^2+f_2\ 32.5^2+\ldots f_8\ 2.5^2+f_{10}\ 7.5^2+f_{12}\ 17.5^2+\ldots\ldots=s$

Then $225=\frac{1}{200}[s+2.5^2f_9+12.5^2f_{11}]$ or $s+6.25f_9+156.25f_{11}$

$=45000$

Let the correct value of the second moment about 40 be δ^2

Then

$$\delta^2=\frac{1}{200}\ [s+6.25\ (f_9+1)+156.25(f_{11}-1)]$$

$$=\frac{1}{200}[s+6.25f_9+156.25f_{11}+6.25-156.25]$$

$$=\frac{1}{200}[45000-150]=\frac{44850}{200}=224.25$$

$\sigma^2=s^2-d^2$ where d is the difference between the actual and assumed mean.

In this example $s^2=\delta^2=224.25$ and $d=(40-39.95)=0.05$

$\therefore\ \sigma^2=224.25-0.0025$

$=224.2475$

$\therefore\ \sigma=14.97$

The corrected standard deviation corresponding to the corrected distribution is 4.97.

PROBLEM 2.48. *The mean age and standard deviation of a group of 100 persons grouped in intervals 10—, 12—, . . etc.) were found to be 32.02 and 13.18. Later it was discovered that the age 57 was misread as 27. Find the corrected mean and tandard deviation.*

SOLUTION. The age 57 belongs to the group 56—58 (mid-value=57) and he age 27 belongs to the group 26—28 (mid-value=27). Let the misread frequenies of these two groups be f_1 and f_2. Then the corrected frequencies will be (f_1+1) nd (f_2-1) respectively. All other frequencies have been entered correctly.

Mid-value	Frequency (wrong)	Frequency (correct)
57	f_1	f_1+1
27	f_2	f_2-1

Value of the mean when 57 was misread as 27 is given by 32.0

$$=\frac{1}{100}(s+57f_1+27f_2)\text{ or }s+57f_1+27f_2=3202$$

Where s is the sum of the products of correct frequencies and values. Let the orrect value of the mean be x, then

$$\overline{x}=\ \frac{1}{100}\left\{s+57(f_1+1)+27\,(f_2-1)\right\}$$

$$= \frac{1}{100}\left\{ s+57f_1+27f_2+57-27 \right\}$$

$$= \frac{1}{100}\left\{ 3202+30 \right\} = \frac{3232}{100} = 32.32$$

$\therefore$ Corrected mean $=32.32$.

Value of the standard deviation when 57 was misread as 27 is given by :

$$(13.18)^2 = \frac{1}{100}\left[s+f_1\,(57-30.02)^2+f_2\,(27-30.02)^2 \right]$$

Where $s=\Sigma f\,(x-\bar{x})^2$ for all correct values of f,

$$\text{or } 174.0124 = \frac{1}{100}(s+727.9204\,f_1+9.1204\,f_2)$$

It is the second moment about 32.02 when 57 was misread as 27. Corre second moment about $32.02=\delta^2$ is given by

$$\delta^2 = \frac{1}{100}(s+727.9204\,(f_1+1)+9.1204\,(f_2-1))$$

$$= \frac{1}{100}(s+727.9204\,f_1+9.1204\,f_2+727.9204-9.1204)$$

$$= \frac{1}{100}(17401.24+718.8) = \frac{18120.04}{100} = 181.2004$$

We know that $\sigma^2=\delta^2-d^2$ where d is the difference between the actual assumed mean.

In this example $s^2=\delta^2=181.2004$, and $d=32.32-32.02=0.3$
$\therefore\ \sigma^2=181.2004-0.09=181.1104$

$\therefore$ Corrrected standard deviation $=\sqrt{181.1104=13.45}$

PROBLEM 2.49. *The mean and the standard deviation of 1000 values variate (grouped in intervals 2.5—7.5,—... etc.) were found to be 29.93 and 9 Later it was discovered that in calculating these values the errata which was supp with the data was not considered. The errata read as follows :*

Group 7.5—12.5 for frequency 30 read frequency 28
Group 17.5—22.5 for frequency 120 read frequency 121
Group 27.5—32.5 for frequency 200 read frequency 198
Group 32.5—37.5 for frequency 175 read frequency 176
Group 47.5—52.5 for frequency 25 read frequency 27

Calculate the corrected mean and standard deviation for the corrected series.

SOLUTION. The frequencies in all other groups were recorded corre Let these frequencies be noted by f_i and let $\Sigma f_i x_i = s$.

The value of the mean when the errata was not considered is given by :

$$29.93 = \frac{1}{1000}\left[s+(30\times 10)+(120\times 20)+(200\times 30)+(175\times 35)+ (25\times 50) \right]$$

$$= \frac{1}{1000}\left[s+300+2400+6000+6125+1250 \right]$$

$$=\frac{1}{1000}[s+16075]$$

$\therefore s=29930-16075=13855$

Correct value of the mean is given by :

$$\bar{x}=\frac{1}{1000}[s+(28\times10)+(121\times20)+(198\times30)+(176\times35)+$$
$$(27\times50)]$$

$$=\frac{1}{1000}[13855+280+2420+5940+6100+1350]$$

$$=\frac{30005}{1000}=30.005$$

Again let $\Sigma fi\,(xi-29.93)^2=T$; where, fi are the correct frequencies. Then second moment about 29.93, when errata was not considered is given by :

$$(9.977)^2=\frac{1}{1000}[T+30\,(10-29.93)^2+120(20-29.93)^2+200$$
$$(30-29.93)^2+175\,(35-29.93)^2+25(50-29.93)^2]$$

$$\text{or } 99.5295=\frac{1}{1000}[T+38318.195]$$

or $T=99529.5-38318.195=61211.305$

Correct second moment about 29.93 is given by

$$\delta^2=\frac{1}{1000}[T+(28\times397.2049)+(121\times98.6049)+(198\times.0049)+$$
$$(176\times25.7049)+(27\times402.8049)]$$

$$=\frac{1}{1000}[61211.305+38453.695]=\frac{1}{1000}\times 99665=99.665$$

We know that $\sigma^2=s^2-d^2$. Here $s^2=\delta^2=99.665$ and $d=.075$

$\therefore \sigma^2=99.665-\text{-}.0056=99.6594$

$\therefore \sigma=\text{-}9.98$

PROBLEM 2.50 *Show that if the deviations are small compared with the mean, so that* $\left(\frac{X}{M}\right)^3$ *and higher powers of* $\frac{X}{M}$ *may be neglected, we have approximately* $G=M\left(1\ \ \frac{1\,\sigma^2}{2M^2}\right)$

Where, G is the Geometric Mean, M the Arithmetic Mean and σ the Standard Deviation : and consequently to the same degree of approximation, $M^2-G^2=\sigma^2$.

SOLUTION Let x_1, x_2......xn be the values of the variate.

The $G=(x_1x_2x_3......xn)\,\frac{1}{n}$

If the deviations from the mean are xi, then $xi=xi-M$ or $xi=M+xi$

$$\therefore G=[(M+x_1)\ (M+x_2)......(M+xn)]\,\frac{1}{n}$$

$$=M\left[\left(1+\frac{x_1}{M}\right)\left(1+\frac{x_2}{M}\right)\ldots\left(1+\frac{x_n}{M}\right)\right]\frac{1}{n}$$

$$=M\left[\left(1+\frac{x_1}{M}\right)\frac{1}{n}\left(1+\frac{x_2}{M}\right)\frac{1}{n}\left(1+\frac{x_n}{M}\right)\frac{1}{n}\right.$$

Expanding the terms on *v. n. s.* and neglecting $\left(\frac{x}{M}\right)^3$ and higher powers when,

$$G=M\left[\left(1+\frac{1}{n}\frac{x_1}{M}+\frac{\frac{1}{n}\left(\frac{1}{n}-1\right)}{2!}\frac{x_1^2}{M^2}\right)\left(1+\frac{1}{n}\frac{x_2}{M}+\right.\right.$$
$$\left.\left.\frac{\frac{1}{n}\left(\frac{1}{n}-1\right)}{2!}\frac{x_2^2}{M^2}\right)\ldots\left(1+\frac{1}{n}\frac{x_n}{M}+\frac{\frac{1}{n}\left(\frac{1}{n}-1\right)}{2!}\frac{x_n^2}{M^2}\right)\right]$$

On multiplying these terms and again ignoring $\left(\frac{x}{M}\right)^3$ and higher powers

$$G=M\left(1+\frac{1}{n\times M}(x_1+x_2+\ldots x_n)+\frac{1-n}{2n^2}\times\frac{1}{M^2}(x_1^2+x_n^2)\right.$$
$$+\frac{1}{n_2M^2}(x_1x_2+x_1x_n+\ldots\ldots+x_{n_1}x_n)$$

Now $x_1+x_2+\ldots+x_n=0$—being the sum of the deviation from the mean, $x_1^2+x_2^2+\ldots+x^2n=n\sigma^2$

and $x_1^2+x_2^2+\ldots x^2n+2x_1x_2+\ldots 2x_{n\ 1}x_n=(x_1+x_2+\ldots+x_n)=0$

$$\therefore G=M\left(1+0-\frac{1}{2n}\times\frac{1}{M^2}(x_1^2+\ldots\ldots+x_n^2)\times\frac{1}{2n^2M^2}(x_1^2+\ldots+x^2n+2x_1x\right.$$
$$\left.\ldots+2x_{n\ 1}x_n)\right)$$

$$=M\left(1-\frac{1}{2n}\times\frac{1}{M^2}\times n\sigma^2\right)=M\left(1-\frac{\sigma^2}{2M^2}\right)$$

(*ii*) $G^2=M^2\left(1-\frac{\sigma^2}{2M^2}\right)^2$

$$=M^2\left(1-\frac{2\sigma^2}{2M^2}\right)\text{—Neglecting }\frac{\sigma^4}{4M^4}$$

or $G^2=M^2-\sigma^2$ or $M^2-G^2=\sigma^2$

PROBLEM 2.51 *If deviations are small compared with the mean, so th* $\left(\frac{x}{M}\right)^3$ *and higher powers of* $\frac{x}{M}$ *may be neglected, show that we have appro mately*

$$H=M\left(1-\frac{\sigma^2}{M^2}\right)$$

Where H is the Harmonic Mean.

SOLUTION. Let the terms of the series be x_1, $x_2\ldots x_n$ and their deviation fr mean M be x_1. $x_2\ldots xn$, so that $xi=M+xi$

$$\text{Now, }\frac{1}{H}=\frac{1}{n}\left\{\frac{1}{x_1}+\frac{1}{x_2}+\ldots\frac{1}{x_n}\right\}$$

$$=\frac{1}{n}\left\{\frac{1}{M+x_3}+\frac{1}{M+x_2}+\ldots+\frac{1}{M+x_n}\right\}$$

$$= \frac{1}{n}\left[\frac{1}{M\left(1+\frac{x_1}{M}\right)} + \frac{1}{M\left(1+\frac{x_2}{M}\right)} + \ldots + \frac{1}{M\left(1+\frac{x_n}{M}\right)}\right]$$

$$= \frac{1}{n} \times \frac{1}{M}\left[\left(1+\frac{x_1}{M}\right)^{-1} + \left(1+\frac{x_2}{M}\right)^{-1} + \ldots + \left(1+\frac{x_2}{M}\right)^{-1}\right]$$

On expanding the terms and neglecting $\left(\frac{x}{M}\right)^2$ and higher powers :

$$\frac{1}{H} = \frac{1}{n} \times \frac{1}{M}\left[\left(1-\frac{x_1}{M}+\frac{x_1^2}{M^2}\right) + \left(1-\frac{x_2}{M}+\frac{x_2^2}{M^2}\right) + \ldots + \left(1-\frac{x_n}{M}+\frac{x_n^2}{M^2}\right)\right]$$

$$= \frac{1}{n} \times \frac{1}{M}\left(n - \frac{1}{M}(x_1 + \ldots x_n) + \frac{1}{M^2}(x_1^2 + \ldots x_n^2)\right)$$

Now $x_1 + x_2 + \ldots\ldots\ldots x_n = 0$ and $x_1^2 + x_2^2 + \ldots x_n^2 = n\sigma^2$

$$\therefore \frac{1}{H} = \frac{1}{n} \times \frac{1}{M}\left(n + \frac{1}{M} \times n\sigma^2\right)$$

$$= \frac{1}{M}\left(1 + \frac{\sigma^2}{M}\right)$$

$$\therefore H = M\left(1 + \frac{\sigma^2}{M^2}\right)$$

$$= \left(1 - \frac{\sigma^2}{M}\right)$$ on neglecting higher powers of $\frac{x}{M}$

REVISIONAL PROBLEMS 2.

2.52. Find Range and its Co-efficient of range from the following :

Marks	No. of Students
0—15	7
15—30	15
30—45	18
45—60	10
60—75	6
75—90	4
	60

[*Ans.* : *Range*=*90*
Co-efficient=*1*]

2.53. Find out middle 50%, middle 80% and Co-efficient of Q. D. from the following table :

Size of Items	Frequency
2	3
4	5
6	10
8	12
10	6
12	4
	40

[*Ans.* : *Q.R.*=*4* ; *Percentile range*=*8*
Co-efficient of Q.D.=*·25*]

2·54. (*a*) Calculate Median and Mean Deviation for the following frequency distribution :

Age (yrs.)	No. of persons	Age (yrs.)	No. of persons
1—5	7	26—30	18
5—10	10	31—35	10
11—15	16	36—40	5
16—20	32	41—45	1
21—25	24		

[*Ans.* : *Median=20* ; *Mean Deviation=6.09*]

(*b*) For a series the mean deviation is 15. Find the most likely value of quartile deviation.

[*Ans.* : *12·5*]

2·55. Using Mean Deviation compare which of the series is more variable

$\left(\text{Use Median and } \frac{N+1}{2}\right)$

A	B	A	B
48,224	2962	54,987	5,682
39,648	1348	42,624	4,222
26,249	8648	32,213	6,498
62,424	7692	72,642	9981
		68,349	3427

[*Ans.* : *B is more variable*]

2·56. Calculate the mean, standard deviation and the Co-efficient of Variation of the grades obtained by 20 students in an examination in statistics :—

62, 85, 73, 81, 74, 58, 66, 72, 54, 84
65, 50, 83, 62, 85, 52, 80, 86, 71, 75

[*Ans.* : $a=79.9$; $\sigma=11.45$; $C.V.=16.1$]

2.57 Calculate standard deviation from the following data :—

Marks	No. of Students
10	8
20	12
30	20
40	10
50	7
60	3

[*Ans.* : $\sigma=13.5$]

2·58. Suppose that a prospective buyer tests bursting pressure of the samples of polythene bags received from two manufacturers A and B. The test reveals the following results :

Bursting pressure (Lbs.)	No. of Bags	
	A	B
5—10	2	9
10—15	9	11
15—20	29	18
20—25	54	32
25—30	11	27
30—35	5	13

Which manufacturer's bags judging from these two samples have the highest average bursting pressure ? Which of these is more uniform in bursting pressure ?

			$\bar{X}$	*C.V.*
Ans. :	*Manf.*	*A*	*21.05*	*23.18*
	,,	*B*	*21.86*	*32.35*

2.59. Lives of two models of refrigerators turned in for new models in a recent survey are :

Life (No. of Years)	Number of Refrigerators	
	Model A	Model B
0— 2	5	2
2— 4	16	7
4— 6	13	12
6— 8	7	19
8—10	5	9
10—12	4	1

What is the average life of each model of these refrigerators ? Which model has more uniformity ?

[*Ans.* : *54.9%*]

2.60. Find out which of the following batsmen is more consistent in scoring. Would you also accept him as a better batsman ? Why ?

Batsman A.	5	7	16	27	39	53	56	61	80	101	105
Batsman B.	0	4	16	21	41	43	57	78	83	90	95

C.V. of Batsman A = *67 08.*

C.V. of Batsman B. = *69.52.*

Since coefficient of variation is less for batsman A., hence batsman A is more consistent. Since the average is also higher in case of batsman A, he is to be regarded as the better batsman also.

2.61. From the following observation prepare a frequency distribution table in ascending order with 5-10 (exclusive method) and find out the mean and the standard deviation.

Scores :	19	18	15	35	32	30	8	12
	17	8	18	19	18	30	36	42
	35	37	30	39	25	24	26	28
	8	17	19	21	24	10	16	15
	18	17	12	21	29	30	31	19
	26	24	24	29	27	25	28	22
	22	21	28	35				

[*Ans.* : = *8.56*]

2.62. From an analysis of monthly wages paid to workers in two organisations C and D, the following results were obtained :

	C	D
No of workers	550	600
Average monthly wages	60	48.5
Variance of the distribution of wages	100	144

Obtain the average monthly wages and the variability in individual wages of all the workers in the two organisations taken together.

[*Ans. Combined Mean* = *Rs. 54*]

[*Combined variability* X_{12}=*Rs. 54*, σ_{12}^2= *Rs 155*

2.63. Calculate coefficient of variation for a series for which the follow results are known :

$$N= 50, \Sigma D= -10, \Sigma D^2=400$$

where D= deviation of items from assumed average 75

(*C. V.*=38

2.64. Mean and standard deviation of 200 items are found to be 60 and 20. at the time of calculations two items are wrongly taken as 3 and 67 instead of 13 a 17, find the correct mean and standard deviation.

(X=*59.8*, σ=*20.0*

2.65. The following are some of the particulars of the distribution of weig of boys and girls in a class :

	Boys	Girls
Number	100	50
Mean Weight	60kg.	45 kgs.
Standard deviation	9	4

Find the mean and standard deviation of the combined data.

(*Ans.* M_{12}=*55*, σ_{12}=*10.*

2.66. The mean and standard deviation of a group of 100 observations we found to be 20 and 3 respectively. After the calculations were made it was fou that three of the observations were incorrect, which were recorded as 21,21 and Find the mean and standard deviation if the incorrect observations are omitted.

(*Correct* X=*20*, *correct* σ=*3*

2.67. (*i*) Coefficient of variation of two series are 58 and 69. Their standa deviations are 21.2 and 15.6 respectively. What are their arithmetic means ?

(*Ans.* M = *36.5 and*

(*ii*) Mean of 100 items is 50 and their standard deviation is 4. Find the su of squares of all items.

(*Ans.* ΣX^2 = *160*

2.68. Mean and standard deviations of two items are found to be 40 and 1 If at the time of calculations two items are wrongly taken as 30 and 70 instead of and 27, find the correct mean and correct standard deviation.

(*Ans. Thus the correct mean* is 39.3 *and the correct standard deviation* 10.24)

2.69. Two samples of size 40 and 50 respectively have the same mean 53, b adifferent standard deviations 19 and 8 respectively. Find the standard deviation the combined sample size of 90.

(*Ans* : σ_{12}=

Moments, Kurtosis and Skewness

3

SKEWNESS

Skewness denotes the absence of symmetry and is indicated where the values of Mean, Median and Mode are not identical. It is calculated by various methods which are illustrated by the examples given in this section.

PROBLEM 3.1. *Find the Second Moment of Dispersion and a co-efficient of Skewness from the data in the following series :—*

Size of item	*Frequency*
3.5	*3*
4.5	*7*
5.5	*22*
6.5	*60*
7.5	*85*
8.5	*32*
9.5	*8*

SOLUTION. Calculation of the Second Moment of Dispersion.

Size of item	Frequency	Deviation from as. av. (6.5)	Square of deviation	Frequency × deviation	Frequency × square of deviation
(m)	(f)	dx	dx^2	fdx	fdx^2
3.5	3	-3	9	-9	27
4.5	7	-2	4	-14	28
5.5	22	-1	1	-22	22
6.5	60	0	0	0	0
7.5	85	$+1$	1	$+85$	85
8.5	32	$+2$	4	$+64$	128
9.5	8	$+3$	9	$+24$	72
	$n=217$			$\Sigma fdx = +128$	$\Sigma fd^2x = 362$

Arithmetic average or $a = x + \dfrac{\Sigma fdx}{n} = 6.5 + \dfrac{128}{217} = 7.089$

$$\text{Second Moment of Dispersion} = \frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2$$

$$= \frac{362}{217} - \left(\frac{128}{217}\right)^2$$

$$= 1.668 - .347$$

$$= 1.321$$

$$\text{Standard Deviation} = \sqrt{\text{Second moment}}$$

$$= \sqrt{1.32}$$

$$= 1.149$$

Location of Mode by grouping

Size of item	Frequency					
	(1)	(2)	(3)	(4)	(5)	(6)
3.5	3	10		32		
4.5	7		29		89	
5.5	22	82				167
6.5	60		145	177		
7 5	85	117			125	
8.5	32		40			
9.5	8					

Analysis Table

Column	Size of item containing maximum frequency				
(1)			7.5		
(2)			7.5	8.5	
(3)		6.5	7.5		
(4)		6.5	7.5	8.5	
(5)			7.5	8.5	9.5
(6)	5.5	6.5	7 5		
No. of times a size occurs	1	3	6	3	1

The size 7.5 is therefore the mode.

Applying to the above data the Karl Pearson's formula of skewness, *viz.* $j=\frac{a-z}{\sigma}$; where j represents the co-efficients of skewness ; *a*, the arithmetic mean ; *z*, the mode and σ, the standard deviation,

We get,
$$j=\frac{7.089-7.5}{1.149}$$
$$=-.359$$

Thus, the second moment of dispersion=1.32 and the co-efficient of skewness=−.359

PROBLEM 3.2. *Find out the mean wage and a co-efficient of skewness for the following—*

35 men get at the rate of Rs. 4.50 per man
40 men get at the rate of Rs. 5.50 per man
48 men get at the rate of Rs. 6.50 per man
100 men get at the rate of Rs. 7.50 per man
125 men get at the rate of Rs. 8.50 per man
87 men get at the rate of Rs. 9.50 per man
43 men get at the rate of Rs. 10.50 per man
22 men get at the rate of Rs. 11.50 per man

SOLUTION. Calculation of Standard Deviation.

Wages in Rs. (m)	No. of men (f)	Deviation from the assumed average(7.5) (dx)	Square of Deviation (d^2x)	Frequency × Deviation fdx	Frequency × Square of Deviation (fd^2x)
4.5	35	−3	9	−105	315
5.5	40	−2	4	− 80	160
6.5	48	−1	1	− 48	48
7.5	100	0	0	0	0
8.5	125	+1	1	+125	125
9.5	87	+2	4	+174	348
10.5	43	+3	9	+129	387
11.5	22	+4	16	+ 88	352
	$n=500$			$\Sigma fdx = +283$	$\Sigma fd^2x = 1735$

Mean wage or $a = x + \frac{\Sigma fdx}{n} = 7.5 + \frac{283}{500} =$ Rs. 8.066

Standard Deviation or $\sigma = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2}$

$$= \sqrt{\frac{1735}{500} - \left(\frac{283}{500}\right)^2}$$

$$= \sqrt{3.15}$$

$= 1.775$ rupees.

Location of mode by grouping

Wages in Rs.	No. of men (f)					
(m)	(1)	(2)	(3)	(4)	(5)	(6)
4.5	35	75		123		
5.5	40		188		188	
6.5	48	148				273
7.5	100		225	312		
8.5	125	212			255	
9.5	87		130			152
10.5	43	65				
11.5	22					

Analysis Table

Column	Size of item containing maximum frequency				
(1)			8.5		
(2)			8.5	9.5	
(3)		7.5	8.5		
(4)		7.5	8.5	9.5	
(5)			8.5	9.5	10.5
(6)	6.5	7.5	8.5		
No. of times a size occurs	1	3	6	3	1

The wages Rs. 8.5 is, therefore, the modal wage.

Applying the above data to the formula.

We get, $$j=\frac{a—z}{\sigma}$$

We get, $$j=\frac{8.066—8.5}{1.77}$$

$$=\frac{—.434}{1.77}$$

$$=—.2451$$

Thus the mean wage=Rs. 8.066.

and Co-efficient of skewness=—.2451.

PROBLEM. 3.3. *Find the co-efficient of skewness from the following data :—*

	Heights of school boys				
Height in centimetres	*No.*	*Height in centimetres*	*No.*	*Height in centimetres*	*No.*
128	*1*	*136*	*166*	*144*	*567*
129	*0*	*137*	*344*	*145*	*233*
130	*1*	*138*	*740*	*146*	*89*
131	*3*	*139*	*1167*	*147*	*27*
132	*8*	*140*	*1670*	*148*	*19*
133	*13*	*141*	*1614*	*149*	*4*
134	*40*	*142*	*1541*	*150*	*4*
135	*59*	*143*	*1028*	*151*	*1*

SOLUTION. Calculation of the Standard Deviation

Height in centimetres (m)	No. of boys (f)	Deviation from assumed average (140) dx	Square of deviation d^2x	Frequency × Deviation fdx	Frequency × Square of deviation fd^2x
128	1	—12	144	— 12	144
129	0	—11	121	— 0	0
130	1	—10	100	— 10	100
131	3	— 9	81	— 27	243
132	8	— 8	64	— 64	512
133	13	— 7	49	— 91	637
134	40	— 6	36	— 240	1440
135	59	— 5	25	— 295	1475
136	166	— 4	16	— 664	2656
137	344	— 3	9	—1032	3096
138	740	— 2	4	—1480	2960
139	1167	— 1	1	—1167	1167
140	1670	0	0	0	0
141	1614	+ 1	1	+1614	1614
142	1541	+ 2	4	+3082	6164
143	1028	+ 3	9	+3084	9252
144	567	+ 4	16	+2268	9072
145	233	+ 5	25	+1165	5825
146	89	+ 6	36	+ 534	3204
147	27	+ 7	49	+ 189	1323
148	19	+ 8	64	+ 152	1216
149	4	+ 9	81	+ 36	324
150	4	+10	100	+ 40	400
151	1	+11	121	+ 11	121
	$n=9339$			$\Sigma fdx=7093$	$\Sigma fd^2x=52945$

Mean height or $a = x + \frac{\Sigma fdx}{n} = 140 + \frac{7093}{9339} = 40.75$ centimetres

Standard Deviation of the height or $\sigma = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2}$

$$= \sqrt{\frac{52945}{9339} - \left(\frac{7093}{9339}\right)^2}$$

$$= \sqrt{5.67 - .56} = \sqrt{5.11}$$

$$= 2.261. \text{ centimetres}$$

Location of Mode by grouping

Height in centimetres (*m*)	No. (*f*) (1)	(2)	(3)	(4)	(5)	(6)
128	1	1		2		
129	0		1		4	
130	1	4				12
131	3		11	24		
132	8	21			61	
133	13		53			112
134	40	99		265		
135	59		225		569	
136	166	510				1250
137	344		1084	2251		
138	740	1907			3577	
139	1167		2837			**4451**
140	1670	**3284**		**4825**		
141	1614		3155		**4183**	
142	1541	2569				3136
143	1028		1595	1828		
144	567	800			889	
145	233		322			349
146	89	116		135		
147	27		46		50	
148	19	23				27
149	4		8	9		
150	4	5				
151	1					

Analysis Table

Column	Size of item containing maximum frequency				
(1)		140			
(2)		140	141		
(3)			141	142	
(4)		140	141	142	
(5)			141	142	143
(6)	139	140	141		
No. of times a size occurs	1	4	5	3	1

The modal height or z is therefore 141 centimetres

Applying to the above data the Karl Pearson's formula,

$$j=\frac{a-z}{\sigma}$$; where j represents co-efficient of skewness

We have

$$j=\frac{40.76-41.00}{2.26}$$

$$=\frac{-.24}{2.26}$$

$$=-.1061$$

Thus the required co-efficient of skewness is — . 1061.

PROBLEM. 3.4. *Find out co-efficient of dispersion and a co-efficient of skewness from the following table giving wages of 23 persons and explain their significance.*

Wages in Rs.	*No. of persons*	*Wages in Rs.*	*No. of persons*
270—280	*12*	*310—320*	*50*
280—290	*18*	*320—330*	*45*
290—300	*35*	*330—340*	*20*
300—310	*42*	*340—350*	*8*

SOLUTION. Calculation of Standard Deviation.

Wages in Rs. (m)	No. of persons (f)	Mid values ($m.v.$)	Deviation from assumed average (305) (dx)		Square of Deviation d^2x	Frequency × Deviation fdx	Frequency × square of deviation fd^2x
				dx/i where $i=10$			
270—280	12	275	—30	—3	9	—36	108
280—290	18	285	—20	—2	4	—36	72
290—300	35	295	—10	—1	1	—35	35
300—310	42	305	0	0	0	0	0
310—320	50	315	+10	+1	1	+50	50
320—330	45	325	+20	+2	4	+90	180
330—340	20	335	+30	+3	9	+60	180
340—350	8	345	+40	+4	16	+32	128
	$n=230$					$\Sigma fdx=+125$	$\Sigma fd^2x=753$

Arithmetic average, or $a=x+\dfrac{\Sigma fdx}{n}\times i=305+\dfrac{125}{230}\times 10$

$=310.43$ rupees.

$$\text{Standard Deviation}=\sqrt{\frac{\Sigma fd^2x}{n}-\left(\frac{fdx}{n}\right)^2}\times i$$

$$=\sqrt{\frac{753}{230}-\left(\frac{125}{230}\right)^2}\times 10$$

$$=\sqrt{2.978}\times 10$$

$$=1.725\times 10$$

$=17.25$ rupees.

Location of Mode by grouping

Wages in Rs. (m)	No. of persons (frequency) (1)	(2)	(3)	(4)	(5)	(6)
270—280	12	30		65		
280—290	18		53		95	
290—300	35	77				127
300—310	42		92	137		
310—320	50	95			115	
320—330	45		65			73
330—340	20	28				
340—350	8					

Analysis Table

Column	Size of item containing maximum frequency				
(1)			310—320		
(2)			310—320	320—330	
(3)		300—310	310—320		
(4)		300—310	310—320	320—330	
(5)			310—320	320—330	330—340
(6)	290—300	300—310	310—320		
No. of times a size occurs	1	3	6	3	1

The modal group is, therefore, the Rs. 310—320 wage group

By interpolation,

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1) = 310 + \frac{50 - 42}{100 - 42 - 45}(320 - 310)$$

$$= 316.15$$

Thus the modal wages (Z) is Rs. 316.15

Applying to the above data Karl Pearson's formula.

$j = \frac{a - z}{6}$; where j represents the co-efficient of skewness,

We get,

$$j = \frac{310.43 - 316.15}{17.25} = \frac{-5.72}{17.25} = -.332.$$

Thus the required co-efficient of skewness is —.332.

PROBLEM. 3·5. *The following table gives the distribution of population in towns A and B in age groups. Compare the variation and skewness of their frequencies.*

Age-group	*Population in thousands*	
	A	*B*
0—10	*18*	*10*
10—20	*16*	*12*
20—30	*15*	*24*
30—40	*12*	*32*
40—50	*10*	*29*
50—60	*5*	*11*
60—70	*2*	*3*
above 70	*1*	*1*

Calculation of Standard Deviation

					A			B	
Age group (*m*)	M. value (*m.v.*)	Deviation from as. av. (25) (dx)	Square of deviation (d^2x)	Population in thousands (f)	Frequency × Deviation (fdx)	Frequency × square of Deviation (fd^2x)	Population in thousands (f)	Frequency × Deviation (fdx)	Frequency × square of Deviation (fd^2x)
0—10	5	—20	400	18	—360	7200	10	—200	4000
10—20	15	—10	100	16	—160	1600	12	—120	1200
20—30	25	0	0	15	0	0	24	0	0
30—40	35	+10	100	12	+120	1200	32	+320	3200
40—50	45	+20	400	10	+200	4000	29	+580	11600
50—60	55	+30	900	5	+150	4500	11	+330	9900
60—70	65	+40	1600	2	+ 80	3200	3	+120	4800
70—80	75	+50	2500	1	+ 50	2500	1	+ 50	2500
				$n=79$	$\Sigma fdx=+80$	$\Sigma fd^2x=$ 24,200	$n=122$	$\Sigma fdx=$ +1,080	$\Sigma fd^2x=$ 37,200

Place A

Average Age of population or $a = x + \frac{\Sigma fdx}{n} = 25 + \frac{80}{79}$

$= 26$ years

Standard Deviation or $\sigma = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2}$

$= \sqrt{\frac{24200}{79} - \left(\frac{80}{79}\right)^2} = \sqrt{305.32}$

$= 17.5$ years

Co-efficient of variation $= \frac{\sigma}{a} \times 100 = \frac{17.5}{26} \times 100 = 67.3$

Place B

Average Age of population or $a = x + \frac{\Sigma fdx}{n}$

$= 25 + \frac{1080}{122}$

$= 33.85$ years

Standard Deviation or $\sigma = \sqrt{\frac{\Sigma fdx}{n} - \left(\frac{\Sigma fdx}{n}\right)^2}$

$= \sqrt{\frac{37200}{122} - \left(\frac{1080}{122}\right)^2} = \sqrt{226.6}$

$= 15.05$ years

Co-efficient of Variation: $\frac{\sigma}{a} \times 100 = \frac{15.05 \times 100}{33.85} = 44.5$

Location of Mode by grouping

Population in thousands

Age-group	Place A (Frequency)						Place B (Frequency)					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
0—10	18	34		49			10	22		46		
10—20	16		31		43		12		36		68	
20—30	15	27				37	24	56				85
30—40	12		22	27			32		61	72		
40—50	10	15			17		29	40			43	
50—60	5		7			8	11		14			15
60—70	2	3					3	4				
70—80	1						1					

Analysis Table

Size of item containing maximum frequency

Column	Place A					Place B				
(1)	0-10							30-40		
(2)	0-10	10-20					20-30	30-40		
(3)		10-20	20-30					30-40	40-50	
(4)	0-10	10-20	20-30					30-40	40-50	
(5)		10-20	20-30	30-40		10-20	20-30	30-40		50-60
(6)			20-30	30-40	40-50		20-30	30-40	40-50	
No. of times a size occurs	3	4	4	2	1	1	3	6	3	1

As mode is ill defined, the median has to be calculated

Computation of median

Age-group	Population in thousands			
	Place A		Place B	
	Frequency	Cumulative Frequency	Frequency	Cumulative Frequency
0-10	18	18	10	10
10-20	16	34	12	22
20-30	15	49	14	46
30-40	12	61	32	78
40-50	10	71	29	107
50-60	5	76	11	118
60-70	2	78	3	121
70-80	1	79	1	122

Place A

Median age = the age of the $\frac{79}{2}$ *i.e.* 39.5 thousandth person which lies in 20 – 30 age group,

By interpolation

$$l_1 + \frac{l_2 - l_1}{f_1}(m - c)$$

$$= 20 + \frac{30 - 20}{15}(39.5 - 34) = 23.67$$

= 24 years approx.

Place B

Median age = the age of the $\frac{122}{2}$ *i.e.* 61 thousandth person

Which lies in 30 – 40 age group.

By interpolation

$$=30+\frac{40-30}{32}(61-46)=34.69$$

$=35$ years approx.

Applying to the above data Karl Pearson's formula.

$j=\frac{3(a—M)}{\sigma}$; where j represents the co-efficient of skewness ; and a, M, and σ, the average, the Median and the Standard Deviation respectively.

we get, j (of place A)$=\frac{3(26—24)}{17.5}=\frac{6}{17.5}=+.3428.$

j (of place B)$=\frac{3(33.85—35)}{15.05}=\frac{—3.45}{15.05}=—.2292.$

Thus, on comparison of co-efficients of variation and skewness it is obvious that variation and skewness are greater in A.

PROBLEM 3.6. *From the following table state in which section the co-efficient of skewness is higher.*

Marks obtained	*Students of sec. A*	*Students of sec. B.*
0—10	*1*	*0*
10—20	*4*	*5*
20—30	*10*	*10*
30—40	*22*	*13*
40—50	*30*	*42*
50—60	*35*	*30*
60—70	*10*	*10*
70—80	*7*	*8*
80—90	*1*	*2*

SOLUTION. Calculation of standard deviation.

					Section A			Section B		
Marks	M. values	Deviations from as av. (45)		Square of deviation	No, of students	Frequency × deviation	Frequency × square of deviation	No. of students	Frequency × deviation	Frequency × square of deviation
(m)	(m.v.)	(dx)		(d^2x)	(f)	(fdx)	(fd^2x)	(f)	(fdx)	fd^2x)
		dx^2i where $i=10$								
0—10	5	—40	—4	16	1	— 4	16	0	0	0
10—20	15	—30	—3	9	4	—12	36	5	—15	45
20—30	25	—20	—2	4	10	—20	40	10	—20	40
30—40	35	—10	—1	1	22	—22	22	13	—13	13
40—50	45	0	0	0	30	0	0	42	0	0
50—60	55	+10	+1	1	35	+35	35	30	+30	30
60—70	65	+20	+2	4	10	+20	40	10	+20	40
70—80	75	+30	+3	9	7	+21	63	8	+24	72
80—90	85	+40	+4	16	1	+ 4	16	2	+ 8	32
					$n=$ 120	$\Sigma fdx=$ +22	Σfd^2x =268	$n=$ 120	$\Sigma fdx=$ +34	Σfd^2x = 272

Series A

$$\text{Arithmetic average or } a = x + \left(\frac{\Sigma fdx}{n} \times i\right) = 45 + \left(\frac{22}{120}\right) \times 10$$

$$= 46.83 \text{ marks.}$$

$$\text{Standard Deviation or } \sigma_1 = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i$$

$$= \sqrt{\frac{268}{120} - \left(\frac{22}{120}\right)^2} \times 10$$

$$= \sqrt{2.2} \times 10$$

$$= 1.483 \times 10$$

$$= 14.83 \text{ marks}$$

Section B

$$\text{Arithmetic average or } a = x + \left(\frac{\Sigma fdx}{n}\right) \times i = 45 + \left(\frac{34}{120}\right) \times 10$$

$$= 47.83 \text{ marks}$$

$$\text{Standard Deviation or } \sigma_2 = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i$$

$$= \sqrt{\frac{272}{120} - \left(\frac{34}{120}\right)^2} \times 10$$

$$= \sqrt{2.185} \times 10$$

$$= 1.478 \times 10$$

$$= 14.78 \text{ marks.}$$

Location of Mode by grouping

Marks (*m*)	No. of students (Frequency)											
	Section A						Section B					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
0—10	1	5		15			0	5		15		
10—20	4		14		36		5		15		28	
20—30	10	32				62	10	23				65
30—40	22		52	87			13		55	85		
40—50	30	65			75		42	72			82	
50—60	35		45			52	30		40			48
60—70	10	17		18			10	18		20		
70—80	7		8				8		10			
80—90	1						2					

Analysis Table

Column	Size of item containing maximum frequency									
	Section A					Section B				
(1)				50-60				40-50		
(2)			40-50	50-60				40-50	50-60	
(3)				50-60	60-70		30-40	40-50		
(4)		30-40	40-50	50-60			30-40	40-50	50-60	
(5)			40-50	50-60	60-70			40-50	50-60	60-70
(6)	20-30	30-40	40-50			20-30	30-40	40-50		
No. of times a size occurs	1	2	4	5	2	1	3	6	3	1

The Modal marks group of Section A is the 50—60 marks group and Modal marks group of Section B is the 40—50 marks group

By Interpolation

Modal marks of section A or $Z_a = l_1 + \frac{f_1-f_0}{2f_1-f_0-f_2}(l_2-l_1)$

$= 50 + \frac{35-30}{70-30-10}(60-50)$

$= 51.67$ marks approx.

Modal marks of Section B or $Z_b = 40 + \frac{42-13}{84-13-30}(50-40)$

$= 47.07$ marks approx.

Applying to the above data Karl Pearson's formula.

$j = \frac{a-z}{a}$, where j represents the co-efficient of skewness.

We have,

j (of Section A) $= \frac{46.83-51.67}{14.8} = -.327.$

j (of Section B) $= \frac{47.83-47.07}{14.8} = +.051.$

Thus the co-efficient of skewness is higher in Section A.

PROBLEM. 3.7. *Weekly earnings of two groups of workers in factory A and B are given and you are asked to compute co-efficient of skewness by Pearson's method.*

Weekly Wages Rs.	*No. of workers of Factory A*	*No. of workers of Factory B*
40—50	*5*	*30*
50—60	*6*	*35*
60—70	*8*	*40*
70—80	*10*	*50*
80—90	*25*	*70*
90—100	*30*	*35*
100—110	*46*	*15*
110—120	*50*	*13*
120—130	*60*	*12*
130—140	*70*	*10*
	Total 310	*310*

SOLUTION. Calculation of the standard deviation of the weekly earnings of two groups of workers in Factories A and B

Weekly wage	M. Values	Deviation from as. av. (95).	Taking the common (10) from deviations	Square of deviation	Factory A			Factory B		
					No. of workers	Frequency × deviation	Frequency × square of deviation	No. of workers	Frequency × deviation	Frquency × square of deviation
(m)	$(m.v.)$	(d)	(dx)	(d^2x)	(f)	(fdx)	(fd^2x)	(f)	(fdx)	(fd^2x)
40— 50	45	—50	—5	25	5	— 25	125	30	—150	750
50— 60	55	—40	—4	16	6	— 24	96	35	—140	560
60— 70	65	—30	—3	9	8	— 24	72	40	—120	360
70— 80	75	—20	—2	4	10	— 20	40	50	—100	200
80— 90	85	—10	—1	1	25	— 25	25	70	— 70	70
90—100	95	0	0	0	30	0	0	35	0	0
100—110	105	+10	+1	1	46	+ 46	46	15	+ 15	15
110—120	115	+20	+2	4	50	+100	200	13	+ 26	52
120—130	125	+30	+3	9	60	+180	540	12	+ 36	108
130—140	135	+40	+4	16	70	+280	1120	10	+ 40	160
					$n = 310$	+488 $= \Sigma fdx$	2264 $= \Sigma fd^2x$	$n = 310$	—463 $= \Sigma fdx$	2275 $= \Sigma fd^2x$

Factory A

$$\text{Mean of the weekly wages} = x + \frac{\Sigma fdx}{n} \times i$$

$$= 95 + \frac{488}{310} \times 10$$

$$= 110.74 \text{ rupees.}$$

$$\text{Standard Deviation or } \sigma_1 = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i$$

$$= \sqrt{\frac{2264}{310} - \left(\frac{488}{310}\right)^2} \times 10$$

$$= \sqrt{4.85} \times 10$$

$$= 2.202 \times 10$$

$$= 22.02 \text{ rupees.}$$

Factory B

$$\text{Mean of the weekly wages} = x + \frac{\Sigma fdx}{n} \times i$$

$$= 95 + \frac{-463}{310} \times 10$$

$$= 80.07 \text{ rupees}$$

$$\text{Standard Deviation or } \sigma_2 = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i$$

$$= \sqrt{\frac{2275}{310} - \left(\frac{-463}{310}\right)^2} \times 10$$

$$= \sqrt{5.11} \times 10$$

$$= 2.261 \times 10$$

$$= 22.61 \text{ rupees.}$$

Location of Mode by grouping

Weekly wages in Rs.	Number of workers (frequency)											
	Factory A						Factory B					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
40—50	5	11		19			30	65		105		
50—60	6		14		24		35		75		125	
60—70	8	18				43	40	90				160
70—80	10		35	65			50		120	155		
80—90	25	55			101		70	105			120	
90—100	30		76			126	35		50			63
100—110	46	96		156			15	28		40		
110—120	50		110		180		13		25		35	
120—130	60	130					12	22				
130—140	70						10					

Analysis Table

Column	Size of item containing maximum frequency									
	Factory A					Factory B				
(1)					130-140				80-90	
(2)				120-130	130-140				80-90	90-100
(3)			110-120	120-130				70-80	80-90	
(4)		100-110	110-120	120-130				70-80	80-90	90-100
(5)			110-120	120-130	130-140	50-60	60-70	70-80		
(6)	90-100	100-110	110-120				60-70	70-80	80-90	
No. of times a size occurs	1	2	4	4	3	1	2	4	5	2

As mode is ill-defined in case of Factory A, Median has to be calculated for the purpose of calculating the co-efficient of skewness.

Calculating the Median Wages

Weekly wages in Rs.	No. of workers in Factory A		No. of workers in Factory B	
	Frequency	Cumulative Frequency	Frequency	Cumulative Frequency
40— 50	5	5	30	30
50— 60	6	11	35	65
60— 70	8	19	40	105
70— 80	10	29	50	155
80— 90	25	54	70	225
90—100	30	84	35	260
100—110	46	130	15	275
110—120	50	180	13	288
120—130	60	240	12	300
130—140	70	310	10	310

Factory A

Median wages = the wages of the $\frac{310}{2}$ *i. e.* 155th worker

which lies in Rs. 110—120 wages group,

By interpolation

$$= l_1 + \frac{l_2 - l_1}{f_1}(m - c)$$

$$= 110 + \frac{120 - 110}{50}(155 - 130)$$

= 115 rupees.

Factory B

Median wages = the wages of the $\frac{310}{2}$ *i. e.* 155th worker

which lies in Rs. 80—90 wage group

By interpolation

$$= 80 + \frac{90 - 80}{70}(155 - 155)$$

= 80.14 rupees.

Applying to the above data, Karl Pearson's formula of skewness when mod is ill-defined, *viz.*

$$j=\frac{3\,(a-M)}{\sigma}$$, where j represents the co-efficient of skewness,

We get

Co-efficient of skewness of weekly wages in Factory A

$$=\frac{3\,(110.74-115.00)}{22.02}=-.19 \text{ approx.}$$

Co-efficient of skewness of weekly wages in Factory B.

$$=\frac{3\,(80.07-80.14)}{22.61}=-.0093$$

PROBLEM. 3.8. *From the following table compute quartile deviation as wel as the co-efficient of skewness.*

Size	*Frequency*	*Size*	*Frequency*
4—8	*6*	*24—28*	*12*
8—12	*10*	*28—32*	*10*
12—16	*18*	*32—36*	*6*
16—20	*30*	*36—40*	*2*
20—24	*15*		

SOLUTION. Calculation of Quartiles and Median.

Size	Frequency	Cumulative Frequency
4—8	6	6
8—12	10	16
12—16	18	34
16—20	30	64
20—24	15	79
24—28	12	91
28—32	10	101
32—36	6	107
36—40	2	109

(1) Median or M = the size of the $\frac{109}{2}$ *i. e.*, 54.5th item which lies in the 16—20 size group

Applying the formula of interpolation,

$$M=l_1+\frac{l_2-l_1}{f_1}(m-c)$$

We have $M=16+\frac{20-16}{30}(54.5-34)=18.73.$

$= 18.8$ approx.

(2) Lower Quartile or Q_1 = the size of the $\frac{109}{4}$ *i. e.*, 27.25th item which lies in the 12—16 size group

By interpolation, we get

$$Q_1=l_1+\frac{l_2-l_1}{f_1}(q_1-c)$$

$$=12+\frac{16-12}{18}(27.25-16)$$

$$=14.5$$

(3) Upper Quartile or Q_3 = the size of the $\frac{3(109)}{4}$ *i. e.*, 81.75th item

which lies in the 24—28 size group.

Applying the formula of interpolation

$$Q_3 = l_1 + \frac{l_2 - l_1}{f_1}(q_3 - c)$$

We get
$$Q_3 = 24 + \frac{28-24}{12}.(81.75-79)$$
$$= 24.92 \text{ approx.}$$

(4) Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$$= \frac{24.9-14.5}{2} = \frac{10.4}{2}$$
$$= 5.2$$

(5) Co-efficient of Skewness $= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

$$= \frac{24.92 + 14.5 - 2(18.7)}{24.92 + 14.5}$$
$$= \frac{2.02}{10.42}$$
$$= +.193.$$

PROBLEM 3·9. *The weekly wages earned by one hundred workers of a tory are set out in the following table :*

Weekly wages (Rs.)	*No. of workers*	*Weekly wages (Rs.)*	*No. of workers*
32.5-37.5	*12*	*57.5-62.5*	*10*
37.5-42.5	*16*	*62.5-67.5*	*6*
42.5-47.5	*25*	*67.5-72.5*	*3*
47.5-52.5	*14*	*72.5-77.5*	*1*
52.5-57.5	*13*		

(*i*) *Calculate the three quartiles of the above distribution.*

(*ii*) *Find the absolute and relative measures of dispersion and skewness based a these quartiles.*

SOLUTION. Calculation of quartiles and co-efficient of skewness.

Weekly wages (Rs.)	No. of workers (f)	Cumulative frequencies
32.5-37.5	12	12
37.5-42.5	16	28
42.5-47.5	25	53
47.5-52.5	14	67
52.5-57.5	13	80
57.5-62.5	10	90
62.5-67.5	6	96
67.5-72.5	3	99
72.5-77.5	1	100

Q_1=Size of $\left(\frac{n}{4}\right)^{\text{th}}$ item or $\frac{100}{4}$=25th item

Quartile First lies in 37.5 – 42.5 group.

By interpolation

$$Q_1 = 37.5 + \frac{5}{16}(25-12)$$

Applying to the above data, Karl Pearson's formula of skewness when mode is ill-defined, *viz.*

$j = \frac{3(a-M)}{\sigma}$, where j represents the co-efficient of skewness,

We get

Co-efficient of skewness of weekly wages in Factory A

$$= \frac{3(110.74-115.00)}{22.02} = -.19 \text{ approx.}$$

Co-efficient of skewness of weekly wages in Factory B.

$$= \frac{3(80.07-80.14)}{22.61} = -.0093$$

PROBLEM. 3.8. *From the following table compute quartile deviation as well as the co-efficient of skewness.*

Size	*Frequency*	*Size*	*Frequency*
4—8	*6*	*24—28*	*12*
8—12	*10*	*28—32*	*10*
12—16	*18*	*32—36*	*6*
16—20	*30*	*36—40*	*2*
20—24	*15*		

SOLUTION. Calculation of Quartiles and Median.

Size	Frequency	Cumulative Frequency
4—8	6	6
8—12	10	16
12—16	18	34
16—20	30	64
20—24	15	79
24—28	12	91
28—32	10	101
32—36	6	107
36—40	2	109

(1) Median or M = the size of the $\frac{109}{2}$ *i. e.*, 54.5th item

which lies in the 16—20 size group

Applying the formula of interpolation,

$$M = l_1 + \frac{l_2 - l_1}{f_1}(m - c)$$

We have $M = 16 + \frac{20-16}{30}(54.5-34) = 18.73.$

$= 18.8$ approx.

(2) Lower Quartile or Q_1 = the size of the $\frac{109}{4}$ *i. e.*, 27.25th item

which lies in the 12—16 size group

By interpolation, we get

$$Q_1 = l_1 + \frac{l_2 - l_1}{f_1}(q_1 - c)$$

$$= 12 + \frac{16-12}{18}(27.25-16)$$

$$= 14.5$$

(3) Upper Quartile or Q_3 = the size of the $\frac{3(109)}{4}$ *i. e.*, 81.75th item which lies in the 24—28 size group.

Applying the formula of interpolation

$$Q_3 = l_1 + \frac{l_2 - l_1}{f_1}(q_3 - c)$$

We get
$$Q_3 = 24 + \frac{28 - 24}{12}.(81.75 - 79)$$
$$= 24.92 \text{ approx.}$$

(4) Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$$= \frac{24.9 - 14.5}{2} = \frac{10.4}{2}$$
$$= 5.2$$

(5) Co-efficient of Skewness $= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

$$= \frac{24.92 + 14.5 - 2(18.7)}{24.92 + 14.5}$$
$$= \frac{2.02}{10.42}$$
$$= +.193.$$

PROBLEM 3·9. *The weekly wages earned by one hundred workers of a factory are set out in the following table :*

Weekly wages (Rs.)	*No. of workers*	*Weekly wages (Rs.)*	*No. of workers*
32.5-37.5	*12*	*57.5-62.5*	*10*
37.5-42.5	*16*	*62.5-67.5*	*6*
42.5-47.5	*25*	*67.5-72.5*	*3*
47.5-52.5	*14*	*72.5-77.5*	*1*
52.5-57.5	*13*		

(*i*) *Calculate the three quartiles of the above distribution.*

(*ii*) *Find the absolute and relative measures of dispersion and skewness based on these quartiles.*

SOLUTION. Calculation of quartiles and co-efficient of skewness.

Weekly wages (Rs.)	No. of workers (f)	Cumulative frequencies
32.5-37.5	12	12
37.5-42.5	16	28
42.5-47.5	25	53
47.5-52.5	14	67
52.5-57.5	13	80
57.5-62.5	10	90
62.5-67.5	6	96
67.5-72.5	3	99
72.5-77.5	1	100

Q_1=Size of $\left(\frac{n}{4}\right)^{\text{th}}$ item or $\frac{100}{4}$=25th item

Quartile First lies in 37.5 – 42.5 group.

By interpolation

$$Q_1 = 37.5 + \frac{5}{16}(25 - 12)$$

$$=37.5+\frac{5\times 13}{16}=41.56$$

Q_3=Size of $\frac{3(n)}{4}$ th item or $\frac{3(100)}{4}$ =75th item.

Third Quartile lies in 52.5-57.5 group

By interpolation :

$$Q_3=52.5+\frac{5}{13}\ (75-67)$$

$$=52.5+\frac{5\times 8}{13}=55.58$$

Q_2 or Median=Size of $\left(\frac{n}{2}\right)^{\text{th}}$ item or 50th item which lies in 42.5-47 group.

By interpolation :

$$M=42.5+\frac{5}{25}(50-28)$$

$$=42.5+\frac{5\times 22}{25}=46.9$$

Absolute measures of dispersion and skewness

Dispersion :

$$Q.D.=\frac{Q_3-Q_1}{2}=\frac{55.58-41.56}{2}=\frac{14.02}{2}=7.01$$

Co-efficient of Quartile Deviation $=\frac{Q_3-Q_1}{Q_3+Q_1}=\frac{55.58-41.56}{55.58+41.56}=\frac{7.01}{97.14}=.07$

Skewness :

$$SK=Q_3+Q_1-2\text{ median}$$
$$=55.58+41.56-2\times 56.9=3.34$$

Co-efficient of skewness (j) $=\frac{Q_3+Q_1-2M}{Q_3-Q_1}=\frac{55.58+41.56-2\times 46.9}{55.58-41.56}=\frac{3.34}{14.02}=.$

PROBLEM 3·10. *Find the mean deviation and standard deviation of the following table giving the marks obtained by 500 candidates : Calculate a co-efficient of skewness.*

	Number of Marks	No. of Candidates
Below	10	30
,,	20	70
,,	30	120
,,	40	168
,,	50	192
,,	60	354
,,	70	486
,,	80	500

SOLUTION. Calculation of Mean Deviation and Standard Deviation.

Marks group	M values (m)	No. of candidates (f)	Deviation from the as. av. (45) dx	Square of Deviation (d^2x)	Frequency × deviation fdx	Frequency × square of deviation (fd^2x)
0-10	5	30	= 40	1600	− 1200	48000
10-20	15	40	− 30	900	− 1200	36000
20-30	25	50	− 20	400	− 1000	20000
30-40	35	48	− 10	100	− 480	4800
40-50	45	24	0	0	0	0
50-60	55	162	+ 10	100	+ 1620	16200
60-70	65	132	+ 20	400	+ 2640	52800
70-80	75	14	+ 30	900	+ 420	12600
		$n=500$			$\Sigma fdx=800$	$\Sigma fd^2x=190400$

$$\text{Standard Deviation}=\sqrt{\frac{\Sigma fd^2x}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}$$

$$=\sqrt{\frac{190400}{500}-\left(\frac{800}{500}\right)^2}$$

$$=\sqrt{378.24.}$$

$= 19.45$ marks.

$$\text{Arithmetic average}=x+\frac{\Sigma fdx}{n}=45+\frac{800}{500}$$

$= 46.6$ marks.

Deviation from the average (46.6) (+ & − signs ignored) (d)	Frequency (f)	Frequency × Deviation from average (fd)
41.6	30	1248·0
31.6	40	1264·0
21.6	50	1080·0
11.6	48	556.8
1.6	24	38.4
8.4	162	1360.8
18.4	132	2428.8
28.4	14	397.6
		$\Sigma fd=8374.4$

$$\text{Mean Deviation (from Arithmetic average) or } \delta=\frac{\Sigma fd}{n}$$

$$=\frac{8374.4}{500}=16.7488 \text{ marks}$$

Location of Mode by grouping

Marks (*m*)	No. of candidates (*f*) (1)	(2)	(3)	(4)	(5)	(6)
0-10	30	70		120		
10-20	40		90		138	
20-30	50	98				122
30-40	48		72	234		
40-50	24	186			318	
50-60	162		294			308
60-70	132	146				
70-80	14					

Column	Size of item containing maximum frequency				
(1)			50-60		
(2)		40-50	50-60		
(3)			50-60	60-70	
(4)	30-40	40-50	50-60		
(5)		40-50	50-60	60-70	
(6)			50-60	60-70	70-80
No. of items a size occurs	1	3	6	3	1

Thus, the modal marks group is 50 – 60 group.

By interpolation

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}\,(l_2 - l_1)$$

$$= 50 + \frac{162 - 24}{324 - 24 - 132}\;(60—50) = 58.21$$

Thus the modal marks are 58.2.

Applying to the above data Karl Pearson's formula

$j = \frac{a - z}{\sigma}$ where j represents co-efficient of skewness

we have, $j = \frac{46.6 - 58.2}{19.45} = \frac{-11{\cdot}6}{19.45} = -.596$

PROBLEM 3·11. *Find the Standard and Quartile Deviation and a co-efficient of skewness for the following.*

Class	*Frequency*
0- 3	*4*
3- 6	*8*
6-10	*10*
10-12	*14*
12-15	*16*
15-18	*22*

18-20	24
20-24	28
24-25	20
25-30	14
30-32	12
32-36	7

SOLUTION. Calculation of Standard & Quartile Deviations.

Class (m)	M. values (m.v.)	Frequency (f)	Cumulative Frequency (c. f.)	Deviation from the assumed average (21) dx		Square of Deviation d^2x	Frequency × deviation fdx	Frequency × Square of deviation fd^2x
				dx	dx/i			
0- 6	3	12	12	−18	−3	9	−36	108
6-12	9	24	36	−12	−2	4	−48	96
12-18	15	38	74	− 6	−1	1	−38	38
18-24	21	52	126	0	0	0	0	0
24-30	27	34	160	+ 6	+1	1	+34	34
30-36	33	19	179	+12	+2	4	+38	76
		$n=179$					$\Sigma fdx=-50$	$\Sigma df^2x=352$

Arithmetic average or $a = x + \dfrac{\Sigma fdx}{n} \times i = 21 + \dfrac{-50}{179} \times 6$

$= 19.3$ approx.

$$\textit{Standard Deviation or } \sigma = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times \text{i}$$

$$= \sqrt{\frac{352}{179} - \left(\frac{-50}{179}\right)^2} \times 6$$

$$= \sqrt{1.966 - .078} \times 6$$

$$= \sqrt{1.888} \times 6$$

$$= 1.374 \times 6 = 8.244$$

$$= 8.25$$

Lower Quartile or Q_1 = the value of the $\dfrac{179}{4}$ *i.e.*, 44.75th item which lies in the 12 – 18 group

By interpolation

$$Q_1 = l + \frac{l_2 - l_1}{f_1}(q_1 - c)$$

$$= 12 + \frac{13 - 12}{15}(45.75 - 36) = 15.5$$

Upper Quartile or Q_3 = the value of $\dfrac{3(179)}{4}$ *i.e.* 134.25th item which lies in the 24 – 30 size group

Applying the formula of interpolation

$$Q_3 = l_1 + \frac{l_2 - l_1}{f_1}(q_3 - c)$$

We get,

$$Q_3 = 24 + \frac{30-24}{34}\,(134.25-126)$$

$$= 25.5 \text{ approx.}$$

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2} = \dfrac{25.5-15.5}{2} = 5$

Class	Location of mode by grouping					
	Frequency (f)					
(m)	(1)	(2)	(3)	(4)	(5)	(6)
0- 6	12	36		74		
6-12	24		62		114	
12-18	38	90				124
18-24	**52**		86	105		
24-30	34	53				
30-36	19					

Analysis Table

Column	Size of item containing maximum frequency				
(1)			18 – 24		
(2)		12 – 18	18 – 24		
(3)			18 – 24	24 – 30	
(4)			18 – 24	24 – 30	30 – 36
(5)	6 – 12	12 – 18	18 – 24		
(6)		12 – 18	18 – 24	24 – 30	
No. of times a size occurs	1	3	6	3	1

The modal group is, therefore, the 18 – 24 group,
By interpolation

$$\text{Mode or } Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

$$= 18 + \frac{52-38}{104-38-34}\,(24-18) = 20.6 \text{ approx.}$$

Applying the above data to Karl Pearson's following formula of coeffici of skewness,

$j = \dfrac{a - Z}{\sigma}$; where j represents the co-efficient of skewness.

We have,

$$j = \frac{19.3 - 20.6}{8.25} = \frac{-1.3}{8.25} = -.157$$

Thus, the co-efficient of skewness $= -.157$;
the standard deviation $= 8.25$
and the Quartile Deviation $= 5$

PROBLEM 3.12. *Find out the mean deviation, standard deviation and quarti deviation from the following table. Also calculate a co-efficient of skewness.*

Wages in Rs.	*No. of labourers*
above 200	*685*
,, 210	*500*
,, 220	*423*
,, 230	*389*
,, 240	*209*
,, 250	*73*
,, 260	*50*
,, 270	*0*

SOLUTION. Calculation of Mean Deviation, Standard Deviation and Quartile Deviation.

Wages in Rs.	M. values (m)	No. of labourers (f)	Cumulative Frequency (cf)	Deviation from as. av. (235) (dx)		Square of Deviation dx^2	Frequency × Deviation (fdx)	Frequency × Square of Deviation fd^2x	Deviation from average (29) (+ & – Signs ignored) (d)	Frequency × Deviation from av. (fd)
				dx	dx/i					
200—210	205	185	185	—30	−3	9	−555	1665	24	4440
210—220	215	77	262	—20	—2	4	—154	308	14	1078
220—230	225	34	296	—10	—1	1	— 34	34	4	136
230—240	235	180	476	0	0	0	0	0	6	1080
240—250	245	136	612	+10	+1	1	+136	136	16	2176
250—260	255	23	635	+20	+2	4	+ 46	92	26	598
260—270	265	50	685	+30	+3	9	+150	450	36	1800
		n = 685					$\Sigma fdx = -411$	$\Sigma fd^2x = 2685$		$\Sigma fd = 11308$

Arithmetic average or $a = x + \frac{\Sigma fdx}{n} \times i = 235 + \left(\frac{-411}{685} \times 10\right)$

$= \text{Rs. } 229.$

(1) Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma fd^2x}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i$$

$$= \sqrt{\frac{2685}{685} - \left(\frac{-411}{685}\right)^2} \times 10$$

$$= \sqrt{3.56} \times 10$$

$$= 1.87 \times 10 = 18.7 \text{ rupees.}$$

First Quartile or Q_1 = the wages of $\frac{685}{4}$ *i. e.,* 171.25th labourer which lies in Rs. 200—210 wage group.

By interpolation $= 200 + \frac{210-200}{185}(171.25-0)$

$= 209.3$ rupees approx.

Q_3 = the wages of the $\frac{3\,(685)}{4}$ *i.e.,* 513.75 labourer which lies in the Rs. 40—50 wage group.

By interpolation,

$$= 240 + \frac{250-240}{135}(513.75-476)$$

$$= 242.8 \text{ rupees.}$$

(2) Quartile Deviation $= \frac{Q_3-Q_1}{2} = \frac{242.8-209.3}{2} = 16.75$ rupees.

(3) Mean Deviation (from Mean) $= \frac{\Sigma fd}{n} = \frac{11308}{685} = 16.5$ rupees.

Location of Mode by grouping

Wages in Rs.	No. of Labourers (frequency)					
(*m*)	(1)	(2)	(3)	(4)	(5)	(6)
200—210	185	262		296		
210—220	77		111		291	
220—230	34	214				350
230—240	180		316	339		
240—250	136	159			209	
250--260	23					
260—270	50					

Analysis Table

Column	Size of item containing the maximum frequency					
(1)	200—210					
(2)	200—210	210—220				
(3)				230—240	240—250	
(4)				230—240	240—250	250—
(5)		210—220	220—230	230—240		
(6)			220—230	230—240	240—250	
No. of times a size occurs	2	2	2	4	3	1

The Rs. 230—240 wage group is, therefore, the modal wage group. By interpolation.

$$Z = l_1 + \frac{f_1-f_0}{2f_1-f_0-f_2}(l_2-l_1)$$

$$= 230 + \frac{180-34}{360-34-136}(240-230)$$

$$= 237.7 \text{ rupees.}$$

Thus the modal wage is Rs. 237.7

Applying to the above data the formula.

$$j = \frac{a-Z}{\sigma}$$; where j represents co-efficient of skewness

We have, $j = \frac{229-237.7}{18.9} = \frac{-8.7}{18.9} = -.46$

PROBLEM 3.13 *Calculate Karl Pearson's Co-efficient of skewness the following data :—*

Marks		*Number of students*
Above	*0*	*150*
,,	*10*	*140*
,,	*20*	*100*
,,	*30*	*80*
,,	*40*	*80*
,,	*50*	*70*
,,	*60*	*30*
,,	*70*	*14*
,,	*80*	*0*

SOLUTION. Calculation of Karl Pearson's Co-efficient of skewness.

Marks (m)	No. of students (f)	Step Deviation from as. av. (45) (d)	 (d')	(fd')	(fd'^2)	(cf)
0—10	10	+40	−4	−40	160	10
10—20	40	+30	−3	−120	360	50
20—30	20	+20	−2	−40	80	70
30—40	0	+10	−1	0	0	70
40—50	10	0	0	0	0	80
50—60	40	−10	+1	+40	40	120
60—70	16	−20	+2	+32	64	136
70—80	14	−30	+3	+42	126	150
	$n=150$			$\Sigma fd'=-86$	$\Sigma fd'^2=830$	

As the above series is bi-modal, we shall use the following formula given by Karl Pearson for finding out Co-efficient of Skewness.

$$j=\frac{3\,(a-M)}{\sigma}$$

$$a=x+\left(\frac{\Sigma fd'}{n}\times i\right)$$

$$=45+\left(\frac{-86}{150}\times 10\right)$$

$$=39.27 \text{ marks.}$$

$$\sigma=\sqrt{\frac{\Sigma fd'^2}{n}-\left(\frac{\Sigma fd'}{n}\right)^2}\times i$$

$$=\sqrt{\frac{830}{150}-\left(\frac{-86}{150}\right)^2}\times 10$$

$$=22.8 \text{ marks.}$$

Median = The value of $\frac{n}{2}$ or $\frac{150}{2}$ or 75th item.

which lies in 40—50 group.

By interpolation,

$$M=l_1+\frac{l_2-l_1}{f_1}(m-c)$$

$$=40+\frac{50-40}{10}(75-70)$$

$$=40+5$$

$$=45 \text{ marks.}$$

$$j=\frac{3(39.27-45)}{22.8}$$

$$= \frac{17.19}{22.8}$$

$$= -.75$$

It should be noted that in this formula j varies between +3.

PROBLEM. 3.14. *In a certain distribution the following results we obtained :*

$$x\ (mean) = 45.00,\ median = 48.00$$

Coefficient of skewness = —0.4.

The person who gave you the data failed to give the value of standard devia tion and you are required to estimate it with the help of the available information.

SOLUTION.

$$\text{Coefficient of skewness (j)} = \frac{3(\text{Mean—Median})}{\sigma}$$

Substituting the given values, we get

$$j = \frac{3(x - m)}{\sigma}$$

$$-0.4 = \frac{3(45 - 48)}{\sigma}$$

$$-0.4\sigma = -9.$$

$$\sigma = \frac{9}{.4} = 22.5.$$

PROBLEM 3.15. (*a*) *For a moderately skewed data, the arithmetic mean i 100, the co-efficient of variation is 35 and the Karl Pearson's Co-efficient of skew ness is 0.2. Find the mode and median.*

(*b*) *Karl Pearson's co-efficient of skewness of a distribution is +0.40. Its standard deviation is 8 and mean is 30. Find the mode and median of the distribution.*

SOLUTION. (*a*) Co-efficient of Variation (*c.v.*) $= \frac{\sigma}{a} \times 100$

$$35 = \frac{\sigma}{100} \times 100$$

$$\therefore \quad \sigma = 35$$

Calculation of mode :

$$j = \frac{a - Z}{\sigma}$$

$$0.2 = \frac{100 - Z}{35}$$

$$\therefore \quad Z = 93$$

Calculation of median :

$$j = \frac{3(a - m)}{\sigma}$$

$$0.2 = \frac{3\,(100 - m)}{35}$$

$$\therefore \quad M = 97.7 \text{ approx.}$$

(*b*)

$$j = \frac{a - z}{\sigma}$$

$$0.4 = \frac{30 - Z}{8}$$

$$Z = 26.8$$

Mode = 3 Median—2 Mean

26.8 = 3 Median—2 × 30

$\therefore$ Median = 28.93 approx.

The values of various moments about the mean are—

First Moment or $\mu_1 = \frac{1}{n} \Sigma f(x - \bar{x}) = \frac{\Sigma d}{n}$

Second Moment or $\mu_2 = \frac{1}{n} \Sigma f(x - \bar{x}) = \frac{\Sigma d^2}{n}$

Third Moment or $\mu_3 = \frac{1}{n} \Sigma f(x - \bar{x})^3 = \frac{\Sigma d^3}{n}$

and so on.

Where d represents the difference between mean and other values of the series.

Kurtosis is measured by Co-efficient β_2 which is equal to the fourth moment about the mean divided by the square of the second moment about the mean or

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2}$$

The standard value of $\beta_2 = 3$ and this is the value obtained in a Normal Curve. If β_2 is less than 3 the curve is platykurtic and if it is more than 3, leptokurtic.

Skewness denotes the absence of symmetry and is indicated when values of Mean, Median and Mode are not identical. It is studied by various rules which are indicated in the problem solved.

PROBLEM 3.16. *Calculate the first four moments of the following distribution of the mean :*

x :	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
f :	*1*	*8*	*28*	*56*	*70*	*56*	*28*	*8*	*1 = 256*

SOLUTION. First moment about mean $\bar{x} = \mu_1 = \frac{1}{n} \Sigma f(x - \bar{x})$

Second moment about mean $\bar{x} = \mu_2 = \frac{1}{n} \Sigma f(x - \bar{x})^2$

Third moment about mean $\bar{x} = \mu_3 = \frac{1}{n} \Sigma f(x - \bar{x})^3$

Fourth moment about mean $\bar{x} = \mu_4 = \frac{1}{n} \Sigma f(x - \bar{x})^4$

Mean of the above distribution

$$= \frac{(0 \times 1) + (1 \times 8) + (2 \times 28) + (3 \times 56) + (4 \times 70) + (56 \times 5) + (6 \times 28) + (7 \times 8) + (8 \times 1)^2}{256}$$

$$= \frac{8 + 56 + 168 + 280 + 280 + 168 + 56 + 8}{256} = \frac{1024}{256} = 4$$

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

Calculation of Moments.

(1) x	(2) f	(3) $(x—4)$	(4) $f(x—4)$ col.2 × col.3	(5) $f(x—4)^2$ col. × col.4	(6) $f(x—4)^3$ col.3 × col.5	(7) $f(x—4)^4$ col.3 × col
0	1	—4	— 4	16	— 64	256
1	8	—3	— 24	72	—216	648
2	28	—2	— 56	112	—224	448
3	56	—1	— 56	56	— 56	56
4	70	0	--140	0	—560	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	236
	256	0	+140	512	+560	2816

$$\mu_1 = \frac{-140+140}{256} = 0 \qquad \mu_3 = \frac{-560+560}{256} = 0$$

$$\mu_2 = \frac{512}{256} = 2 \qquad \mu_4 = \frac{2816}{256} = 11$$

PROBLEM 3.17. *Express the first four moments about mean in terms of first four moments about an arbitrary origin A.*

SOLUTION. Let the values of the variate be x_1. $x_2 \ldots x_n$ and let the mean be x.

Let $x—x = \bar{x}$ deviation from mean.

$x—A + \zeta$ deviation from arbitrary origin A

and $\bar{x}—A = d$

Then $x—\bar{x} = (x—A)—(\bar{x}—A) = \zeta—d$

Now first moment about mean $= \mu_1 = \dfrac{1}{N}\Sigma f(x—\bar{x}) = \dfrac{1}{N}\Sigma f(\zeta—d)$

$$= \frac{1}{N}\Sigma f - \frac{1}{N}\Sigma fd$$

$$= \frac{1}{N}\Sigma f(x—A)—d$$

$\dfrac{1}{N}\Sigma f(x=A)$ is the first moment about arbitrary origin $A = \delta_1$

$$\therefore \mu_1 = \delta_1 — d = o$$

$$\mu_2 = \frac{1}{N}\cdot\Sigma f(x—\bar{x})^2 = \frac{1}{N}\Sigma f(\zeta—d)^3 = \frac{1}{N}\Sigma f(\zeta^2—2d\zeta + d^2)$$

$$= \frac{1}{N}[\Sigma f\zeta^2—2d\Sigma f\zeta + Nd^2]$$

$$= \frac{1}{N}\Sigma f(x—A)^2—\frac{2d}{N}\Sigma f(\bar{x}-A) + \frac{Nd^2}{N}$$

$$= \frac{1}{N}\Sigma f(x—A)^2—\frac{2d^2}{n} + d^2 = \delta^2—\delta_1{}^2$$

$$\mu_3 = \frac{1}{N}\Sigma f(x—\bar{x})^3 = \frac{1}{N}\Sigma f(\zeta—d)^3 = \frac{1}{N}\Sigma f(\zeta^3—3\zeta^2 d + 3\zeta d^2—d^3)$$

$$= \frac{1}{N}\Sigma f\zeta^3 - \frac{3d^3}{N}\Sigma f\zeta + \frac{3d^2}{N}\Sigma f\zeta - \frac{d^3}{N}\Sigma f$$

$$= \delta_3 - 3\delta_1\,\delta_2 + 3\delta_1{}^2\delta_1 - \delta_1{}^3$$
$$= \delta_3 - 3\delta_1\,\delta_2 + 2\delta_1^3$$

$$\mu_4 = \frac{1}{N}\Sigma f(x-x) = \frac{1}{N}\Sigma f(\zeta-d)^4 = \frac{1}{N}\Sigma f\,(\zeta^4 - 4d\zeta^3 +$$
$$(6d^2\zeta^2 - 4d^3\zeta + d^4)$$

$$= \frac{1}{N}\Sigma f\,\zeta^4 - \frac{4d}{N}\Sigma f\,\zeta^3 + \frac{6d^2}{N}\Sigma f\,\zeta^2$$
$$\frac{4d^3}{N}\Sigma f\,\zeta + \frac{d^4}{N}\Sigma f$$
$$= \delta_4 - 4\delta_1\delta^3 + 6\delta_1{}^2\,\delta_2 - 4\delta_1{}^3 + \delta_1 + \delta_1{}^4$$
$$= \delta^4 - 4\delta_1\delta^3 + 6\delta_1{}^2\delta_2 - 3\delta_1{}^4$$

PROBLEM 3.18. *Calculate the first four moments of the following distribution about $x = 15$ and thence find the moments about the mean of the distribution.*

x:	10	11	12	13	14	15	16	17	18	19	20	21	Total
f:	9	36	75	105	116	107	88	66	45	30	18	5	700

SOLUTION.

(1) x	(2) f	(3) $(x-15)$	(4) $f(x-15)$ col.2 × col.3	(5) $f(x-15)^2$ col.3 × col.4	(6) $f(x-15)^3$ col.3 × col.5	(7) $f(x-15)^4$ col.3 × col.6
10	9	—5	— 45	225	—1125	5625
11	36	—4	—144	576	—2304	9216
12	75	—3	—225	675	—2025	6075
13	105	—2	—210	420	— 840	1680
14	116	—1	—116	116	— 116	116
15	107	0	—740	0	—6410	0
16	88	1	88	88	88	88
17	66	2	132	264	528	1056
18	45	3	135	405	1215	3645
19	30	4	120	480	1920	7680
20	18	5	90	450	2250	11250
21	5	6	30	180	1080	6480
			+595	3879	+7081	52911

Moments about $x = 15$

$$\delta_1 = \frac{-740+595}{700} = \frac{-145}{700} = -0.2071 \qquad \delta_2 = \frac{3879}{700} = 5.5414$$

$$\delta_3 = \frac{-6410+7081}{700} = \frac{671}{700} = 0.9586 \qquad \delta_4 = \frac{52911}{700} = 75.5871$$

Moments about mean :

$$\mu_2 = \delta_2 = \delta_1{}^2 = 5.5414 - (.2071)^2 = 5.5414 - 0.0429 = 5.4985$$

$$\mu_3 = \delta_3 - 3\delta_1\delta_2 + 2\delta_1{}^2 = 0.9586 - 3(-0.2071)(5.5414) + 2(-02071)^2$$
$$= 0.9586 + 3.4429 - 0.0178 = 4.4015 = 0.018$$
$$= 4.3837$$

$$\mu_4 = \delta_4 - 4\delta_1\delta_3 + 6\delta_1{}^2\delta_2 - 3\delta_1{}^4 = 75.5871 - 4(-.2071)(.9586) + 6.$$
$$(.0429)\,(5.5414) - 3(-.2071)^4$$
$$= 75.5871 + 0.7940 + 1.4263 - .0055 = 77.8074 - .0055$$
$$= 77.8019$$

PROBLEM 3.19. *Express the first four moments about an arbitrary origin in terms of moments about mean.*

SOLUTION. Let the values of the variate be $x_1\ x_3 .. x_n$ and let their mean be $\bar{x}$.

Let $\bar{x} - x = x$ deviation from the mean.

$x - A - \zeta$ deviation from arbitrary origin A

and $\bar{x} - A = d$

Then $\zeta = x - A = (x - x) + (x - A) = x + d$

First moment about $A = \frac{1}{N} \Sigma f(x - A) = \frac{1}{N} \Sigma f(x + d)$

$$\mu = \frac{1}{N} \Sigma f x + \frac{d}{N} \Sigma f$$

or $\delta_1 = \mu_1 + d = d$

Second moment about $A = \delta_1 = \frac{1}{N} \Sigma f (x - A)^2 = \frac{1}{N} \Sigma f (x + d)^2$

$$= \frac{1}{N} (\Sigma f x^2 + 2d \Sigma f x + d^2 \Sigma f)$$

$$= \frac{1}{N} \Sigma f x^2 + \frac{d^2 + N}{N}$$

$$= \mu_2 + d^2$$

Third moment about $A = \delta_3 = \frac{1}{N} \Sigma f (x - A)^3 = \frac{1}{N} \Sigma f (x - d)^2$

$$= \frac{1}{N} \Sigma f (x^3 + 3x^2 d + 3dx^3 + a^3)$$

$$= \mu_3 + 3d\,\mu_2 + 3d^2\,\mu_1 + a^3$$

$$= \mu_3 + 3d\,\mu_2 + d^2$$

Fourth moment about $A = \delta_4 = \frac{1}{N} \Sigma f (x - A)^4 = \frac{1}{N} \Sigma f (x - d)^4$

$$= \frac{1}{N} \Sigma f (x^4 + 3x^3 d + 6x^2 d^2 + 4x d^3 + a^4)$$

$$= \mu_4 + 4d\,\mu_3 + 6d^2\,{}_2\mu + 4d^3\,\mu_1 + a^1$$

$$= \mu_4 + 4d\,\mu_3 + 6d^2\,\mu_2 + a^4$$

PROBLEM 3.20. *The first three moments of a distribution about the value 2 of the variable are 1.16 and—40. Show that mean is 3, the variance 15 and $\mu_3 = -86$. Also show that the first three moments about $x = 0$ are 3, 24 and 76.*

SOLUTION. Moments abour $\bar{x} = 2$ are

$\delta_1 = 1$, $\delta_2 = 16$ and $\delta = -40$

or $\delta_1 = \frac{1}{N} \Sigma f (x - 2) = \frac{1}{N} \Sigma f x - \frac{2}{N} \Sigma f = x - 2 = 1$

$\therefore$ $x = 3$ mean $= 3$

Hence $d = \bar{x} - A = 3 - 2 = 1 = \delta_1$

Now, $\mu_2 \delta_2 - \delta_1^2 = 16 - 1 = 15$; variance $= 15$

$u_3 = \delta_2 - 3\,\delta_1 \delta_2 + 2\delta_{13} = -40 - 3.1 \times 16 + 2.1 = -88 + 2 = -86$

To calculate moments about zero, we find

$d = \bar{x} - A = 3 - 0 = 3$

$\therefore$ First moment about zero $= \mu_1 + d = 3 + 3 = 3$

Second moment about zero $= \mu_2 + d^2 = 15 + 9 = 24$

Third moment about zero $= \mu_2 + 3d\mu_2 + d^2$

$= -86 + 3 \times 3 \times 15 + 3^3$

$= -86 + 135 + 27 = 162 - 86$

$= 76$

PROBLEM 3.21. *The first four moments of a distribution about the value 4 are—1.5, 17,—30 and 108 ; find the moments about the mean and the origin.*

SOLUTION. Moments about $x=4$ are

$\delta_1=-1.5,\ \delta_2=17,\ \delta_3=-30$ and $\delta_4=108$

(*ii*) Calculation of mean :

$$\delta_1=\frac{1}{N}\Sigma f(x-4)=\frac{1}{N}\Sigma fx-\frac{4\Sigma f}{N}=x-4=-1.5$$

$\therefore\ x=2.5$

(*ii*) Now $\delta_1=2.5-4=-1.5$

$\mu_2=\delta_2-\delta_1{}^2=17-(-1.5)2=17-2.25=14.75$

$\mu_3=\delta_3-3\,\delta_1\delta_2+2\delta_1{}^3=-30-3\,(-1.5)\,(17)+2\,(-1.5)^3$

$=-30+76.5-6.75=76.5-36.75$

$=39.75$

$\mu_4=\delta_4-4\,\delta_1\,\delta_3+6\,\delta_1{}^2\,\delta_2-3\delta_1{}^4$

$=108-4\,(-1,5)\,(-30)+6\,(-1.5)2\,(17)-3\,(-1.5)^4$

$=108-180+29.5-15.1875=337.5-195.1875$

$=142.3125$

(*iii*) $d=\bar{x}-A=2.5-0=2.5$

1st moment about $0=\mu_1+d=0+2.5=2.5$

2nd moment about $0=\mu_2+d^2=14.75+6.25=21$

3rd moment about $0=\mu_3+3d\mu_2+d^3=39.75+3\,(2.5)\,(14.75)+(2.5)^3$

$=39.75+110.625+15.625$

$=166$

4th moment about $0=\mu_4+4d\mu_3+6d^2\mu_2+a^4$

$=142.3125+4(2.5)\,(39.75)+6(6.25)(14.75)+(2.5)^4$

$=142.3125+397.5+553.125+39.0625$

$=1132$

PROBLEM 3.22. *The first four moments of a distribution about $x=4$ are 1,4, 10 and 45. Show that the mean is 5 and calculate the moments about $x+0$ and about the mean.*

Calculate also β_1 and β_2 and comment on their values.

SOLUTION. First four moments about the arbitrary origin.

$A=4$ are 1, 4, 10 and 45

$$\delta_1=\frac{1}{N}\Sigma f(x-4)=\frac{1}{N}\Sigma xf\frac{4}{N}\Sigma f=x-4=1 \text{ or } x=5$$

(*i*) Moments about the mean.

$\mu_1=0$

$\mu_2=\delta_2-\delta_1{}^2=4-1=3$

$\mu_3=\delta_3-3\delta_1\delta_2+2\delta_1{}^3=10-3\times1\times4+2\times1^3=10-12+2=0$

$\mu_4=\delta_4-4\delta_1\delta_3+6\delta_1{}^2\delta_2-3\delta_1{}^4=45-4\times1\times10\times6+1^2\times4-3.1^4$

$=45-40+24-3=26$

(*ii*) Moments about $x=0$

$\delta_1=x=5=d$

$\delta_2=\mu_2+\delta_1{}^2=3+25=28$

$\delta_3=\mu_3+3\delta_1\delta_2-2\delta_1{}^3=0+3\times5\times28-2\times5^3=420-250=170$

$\delta_4=\mu_4+4\delta_1\delta_2-6\delta_1{}^2\delta_2+3\delta_1{}^4=26\ ,\ 4\times5\times170-6\times25\times28+3\times625$

$=1101$

(*iii*) $\beta_1=\dfrac{\mu_3{}^2}{\mu_2{}^3}=\dfrac{0}{27}=0$ and $\beta_2=\dfrac{\mu_4}{\mu_2{}^2}=\dfrac{26}{9}=2.889$

Since $\beta_1 = 0, \sqrt{\beta_1} = 0$ and therefore the distribution is symmetrical. β_2 is less than 3, therefore the distribution is platycurtic.But $\beta_2 - 3 = -.111$, therefore the Kurtosis is not much.

PROBLEM 3.23. *The first four moments about the working mean 18.5 of a distribution are 0.294, 7.144, 42.409 and 454.980, calculate the moments about the mean.*

Calculate also β_1 *and* β_2, γ_2 *and* γ_3 *and comment upon the skewness and Kurtosis of the distribution.*

SOLUTION. First four moments about the arbitrary origin 28.5 are 0.294, 7.144, 42.409 and 450.980

$$\delta_1 = \frac{1}{N} \Sigma f(x - 28.5) = \frac{1}{N} \Sigma fx - \frac{28.5}{N} = f\bar{x} - 28.5$$

$$= 0.294 \text{ or } \bar{x} = 28.794$$

$$\mu_2 = \delta_2 - \delta_1^2 = 7.144 - (0.294)^2 = 7.144 - 0.086 = 7.058$$

$$\mu_3 = \delta_3 - 3\delta_1\delta_2 + 2\delta_1^2 = 42.409 - 3(0.294)(7.144) + 2(0.294)^3 = 36.151$$

$$\mu_4 = \delta_4 - 4\delta_1\delta_3 + 6\delta_1^2\delta_2 - 3\delta_1^4 = 454.980 - 4(0.294)(42.409) + 6(.294)^2(7.144) - 3(.294)^4 = 408.738$$

$$\text{Now } \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(36.151)^2}{(7.058)^3} = \frac{1306.895}{351.601} = 3.717\ \delta_1$$

$$= \sqrt{\beta_1} = 1.928$$

$$\beta_2 = \frac{\mu_4}{\mu_2{}^2} = \frac{408.738}{(7.058)^2} = \frac{408.738}{49.816} = 8.205,\ \delta_2 = \beta_2 - 3 = 5.205$$

$= \gamma_1 \sqrt{\beta_1} \neq 0$, and indicates considerable skewness.

$\gamma_2 = \beta_2 - 3 \neq 0$, and indicates considerable leptokurtosis.

PROBLEM 3.24. (*a*) *The first four moments of a distribution about* $x = 2$ *are 1,2.5,5.5 and 16. Calculate the four moments about* $\bar{x}$ *and about zero.*

(*b*) *The first four central moments of a distribution are 0,2.5, 0.7 and 18.75. Test the skewness and kurtosis of the distribution.*

SOLUTION. Moments about Mean ($\bar{x}$)

$$\mu_1 = \delta_1 - \delta_1 = 1 - 1 = 0$$
$$\mu_2 = \delta_2 - \delta_1^2 = 2.5 - (1)^2 = 1.5$$
$$\mu_3 = \delta_3 - 3\delta_2\delta_1 + 2\delta_1^3$$
$$= 5.5 - 3(2.5)(1) + 2(1)^3$$
$$= 5.5 - 7.5 + 2 = 0$$
$$\mu_4 = \delta_4 - 4\delta_3\delta_1 + 6\delta_2\delta_1^2 - 3(\delta_1)^4$$
$$= 16 - 4 \times 5.5 \times 1 + 6 \times 2.5 \times 1 - 3(1)^4$$
$$= 16 - 22 + 15 - 3 = 6$$

Moments about zero

δ_1 = assumed mean + first moment about assumed mean = mean

$$\delta_1 = 2 + 1 = 3$$
$$\delta_2 = \mu_2 + \delta_1^2 = 1.5 + (3)^2 = 10.5$$
$$\delta_3 = \mu_3 + 3\delta_2\delta_1 - 2\delta_1^3$$
$$= 0 + 3 \times 10.5 \times 3 - 2(3)^3 = 40.5$$
$$\delta_4 = \mu_4 + 4\delta_1\delta_3 - 6\delta_1^2\delta_2 + 3\delta_1^4$$
$$= 6 + 4(3)(40.5) - 6(3)^2(10.5) + 3(3)^4 = 168.$$

(*b*) For Skewness

We are given

$$\mu_1 = 0,\ \mu_2 = 2.5,\ \mu_3 = .7,\ \mu_4 = 18.75$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(.7)^2}{(2.5)^3} = +0.031$$

Since $\beta_1 = +0.031$, the distribution is slightly skewed, *i.e.*, it is not perfectly symmetrical.

For Testing Kurtosis:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.75}{(2.5)^2} = 3$$

Since β_2 is exactly three, the distribution is mesokurtic.

PROBLEM 3.25. *You are given the following values of higher moments:*

$$\mu_2 = 43.353,\ \mu_3 = -9.774,\ \mu_4 = 5508.567$$

Find the Corrected Values of each one of these, taking into account the class interval which is 3.

SOLUTION.

$$\mu_2\,(\text{Corrected}) = \mu_2\,(\text{Uncorrected}) - \frac{i^2}{12}$$

$$= 43.353 - \frac{(3)^2}{12}$$

$$= 43.353 - .75 = 42.603.$$

$$\mu_4\,(\text{Corrected}) = \mu_4\,(\text{Uncorrected}) - \tfrac{1}{2}\,i^2.\ \mu_2\,(\text{uncorrected}) + \frac{7}{240}\,i^4$$

$$= 5508.567 - \tfrac{1}{2}\,(3)^2\,(43.353) + \frac{7}{240}\,(3)^4$$

$$= 5508.567 - 145.0885 + 2.3625 = 5365.841$$

REVISIONAL PROBLEMS

3.26 *Calculate Karl Pearson's coefficient of skewness from the following data :*

Wages (Rs.)	*No. of workers*	*Wages (Rs.)*	*No. of workers*
10-15	*8*	*30-35*	*62*
15-20	*16*	*35-40*	*32*
20-25	*30*	*40-45*	*15*
25-30	*45*	*45-50*	*6*

(*Ans.* $j = 0.222$)

3.27. *Examine the skewness for the following frequency distribution of blood donors :*

Age (in years)	*Frequency*	*Age (in years)*	*Frequency*
10-19	*3620*	*60-69*	*1200*
20-29	*7045*	*70-79*	*175*
30-39	*9030*	*80-89*	*11*
40-49	*5025*		
50-59	*2710*		

(*Ans.* $SK(J) = +0.14$)

3.28. *From the given data below calculate First quartile, Third quartile and Median and with their help calculate the coefficient of skewness :*

Wages (Rs.)	*No. of workers*	*Wages (Rs.)*	*No. of workers*
12.50-17.50	*12*	*37.50-42.50*	*10*
17.50-22.50	*16*	*42.50-47.50*	*6*
22.50-27.50	*25*	*47.50-52.50*	*3*
27.50-32.50	*17*	*52.50-57.50*	*1*
32.50-37.50	*12*		

(*Ans* : $Q_1 = 21.72$
$Q_3 = 35.05$
$M = 27.1$
Coefficient of skewness = 0.19)

3.29. *From the data given below calculate Karl Pearson's coefficient of skewness and explain its significance :*

Wages (Rs.)	*70-80*	*80-90*	*90-100*	*100-110*	*110-120*	*120-130*	*130-140*	*140-150*
No. of persons	12	18	35	42	50	45	20	8

(*Ans* : $j = -.33$)

3.30. *From the following data calculate the coefficient of skewness based upon the median and the quartiles and comment on your result :*

Age in years	*No. of persons*	*Age in years*	*No. of workers*
Less than 10	20	Less than 60	305
" " 20	65	" " 70	361
" " 30	110	" " 80	380
" " 40	185	" " 90	396
" " 50	230	" " 100	399

(*Ans* : $j = +0.016$)

3.31. (*a*) *A frequency distribution gives the following results :—*

(*i*) *Coefficient of Variation = 5*
(*ii*) *Standard Deviation = 2*
(*iii*) *Karl Pearson's Coefficient of skewness = 0.5*

Find the mode and mean of the distribution

(*Ans* : *mean* = 40, *mode* = 39)

(*b*) *For a distribution Bowley's coefficient of skewness is* -0.36, $Q_1 = 8{\cdot}6$ *and Median* = 12.3. *What is the quartile coefficient of dispersion ?*

(*Ans* : *Coeff. of dispersion* 0·24)

3.32 . *The first three moments of a distribution about the value 3 of the variable are 2, 10 and 30 respectively. Obtain the first three moments about zero. Show also that the variance of the distribution is 6.*

(*Ans* , $V_1 = 5$, $V_2 = 31$, $V_3 = 201$)

3.33. *The first four moments of a distribution are 1, 4, 10 and 46 respectively. Compute the first four central moments and the Beta constant. Comment upon the nature of the distribution.*

(*Ans* : $\mu_1 = 0$, $\mu_1 = 3$, $\mu_1 = 0$, $\mu 4 = 27$, $\beta_1 = 0$, $\beta_2 = 3$)

3.34. *You are given the following information ;*

X	*F*	*X*	*F*
40-41	1	49-50	29
41-42	2	50-51	19
42-43	4	51-52	14
43-44	7	52-53	6
44-45	12	53-54	4
45-46	20	54-55	2
46-47	38		
47-48	52		
48-49	40		

Calculate the first four moments about 47.5. Convert these results into moments about the mean and calculate β_1 *and* β_2.

(*Ans* : $\mu_2 = 5$, $\mu_4 = 114.5$, $\beta_2 = 3.34$)

Correlation

4

Two variables are said to be correlated when the movements or changes in one are followed directly or indirectly by movements or changes in the other and when the greater the magnitude of change in one, the greater it is in the other, also.

In the present chapter we have solved questions on correlation by Karl Pearson's formula, which is in popular use. First few questions deal with a study of correlation in two series of individual observations. After this we have solved questions relating to continuous series or grouped data. Some questions have also been solved by the Rank Correlation method and by the Method of Concurrent Deviations and in the end there are some questions relating to correlation of short time and long time fluctuations.

PROBLEM 4.1. *Find out the correlation co-efficient between heights of father and son from the following data :—*

Height of father in centimetres	*165*	*166*	*167*	*167*	*168*	*169*	*170*	*172*
Height of son in centimetres	*167*	*168*	*165*	*168*	*172*	*172*	*169*	*171*

SOLUTION. Calculation of correlation co-efficient between heights of father and son.

Height of father in centimetres	Deviation from average height (168)	Square of deviation	Height of son in centimetres	Deviation from average height (169)	Square of deviation	Product of deviations
m_1	x	x^2	m_2	y	y^2	xy
165	-3	9	167	-2	4	6
166	-2	4	168	-1	1	2
167	-1	1	165	-4	16	4
167	-1	1	168	-1	1	1
168	0	0	172	$+3$	9	0
169	$+1$	1	172	$+3$	9	3
170	$+2$	4	169	0	0	0
172	$+4$	16	171	$+2$	4	8
Σm_1 1344 $n=8$	0	$\Sigma x^2=36$	$\Sigma m_2=$ 1352, $n=8$	0	$\Sigma y^2=44$	$\Sigma xy=24$

Average height of father $= \frac{\Sigma m_1}{8} = \frac{1344}{8} = 168$

Average height of son $= \frac{\Sigma m_2}{n} = \frac{1352}{8} = 169$

Applying to the above data, the Karl Pearson's formula,

$$r = \frac{\Sigma xy}{n\sigma x\sigma y} = \frac{\Sigma xy}{n\sqrt{\frac{\Sigma x^2}{n}} \times \sqrt{\frac{\Sigma y^2}{n}}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

Where r stands for co-efficient of correlation and other symbols are the same things as in previous chapters.

We get,

$$r = \frac{24}{\sqrt{36 \times 44}} = +.603$$

Thus the correlation co-efficient between heights of father and son is +.603.

The question can be done in the following manner also :—

Calculation of the co-efficient of correlation between the heights of father and son.

Height of father in centimetres (x)	Height of son in centimetres (y)	Square of the height of father (x^2)	Square of the height of son (y^2)	(xy)
165	167	27225	27889	27555
166	168	27556	28224	27888
167	165	27889	27225	27555
167	168	27889	28224	28056
168	172	28224	29584	28896
169	172	28561	29584	29068
170	169	28900	28561	28730
172	171	29584	29241	29412
$Tx=1344$	$Ty=1352$	$\Sigma(x^2)=225828$	$\Sigma(y^2)=228532$	$(xy)=227160$

Substituting the above values in the formula

$$r = \frac{\Sigma(xy) - Tx.\ Ty/N}{\sqrt{(\Sigma(x^2) - Tx^2/N)\Sigma(y^2) - Ty^2/N}}$$

Where r stands for co-efficient of correlation

Σxy stands for the total of the product of item in the two series

Tx and Ty stand for the totals of x and y series respectively

Σx^2 and Σy^2 stand for the total of the squares of item in x and y serie respectively and

N stands for the number of items paired.

$$r = \frac{227160 - 1344 \times 1352/8}{\sqrt{(225828 - (1344)^2/8)(228532 - (1352)^2/8)}}$$

$$= \frac{227160 - 227136}{\sqrt{(225828 - 225792)(228532 - 228488)}}$$

$$= \frac{24}{\sqrt{36 \times 44}}$$

$$= +.603$$

PROBLEM 4.2. *Compute the co-efficient of correlation from the follow data :—*

x	*1200*	*—1000*	*— 800*	*— 400*	*1200*	*1400*	*— 600*	*—10*
y	*—3600*	*3300*	*2400*	*1200*	*—3600*	*—2100*	*1800*	*30*

SOLUTION. Calculation of the co-efficient of correlation between **X** and **Y** series.

X			Y			
Size of items m_1	Deviations from the average (0) (x)	Square of deviations (x^2)	Size of items m_2	Deviations from the aver. (300) (y)	Square of deviations (y^2)	Product of deviations (xy)
1200	1200	1440000	—3600	—3900	15210000	—4680000
—1000	—1000	1000000	3300	3000	9000000	—3000000
— 800	— 800	640000	2400	2100	4410000	—1680000
— 400	— 400	160000	1200	900	810000	— 360000
1200	1200	1440000	—3600	—3900	15210000	—4680000
1400	1400	1960000	—2100	—2400	5760000	—3360000
— 600	— 600	360000	1800	1500	2250000	— 90000
--1000	—1000	1000000	3000	2700	7290000	—2700000
$\Sigma m=0$	$\Sigma x=0$	$\Sigma x^2=$ 80,00,000	$\Sigma m_2=$ 2,400	$\Sigma y=0$	$\Sigma y^2=$ 5,99,40,000	$\Sigma xy=-$ 2,13,60,000

Substituting the above values in the Karl Pearson's formula

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\times\Sigma y^2}}$$ where r represents the co-efficient of correlation

We get,

$$r=\frac{-21360000}{\sqrt{8000000\times59940000}}$$

$$=\frac{-21360000}{10000\sqrt{800\times5994}}$$

$$=\frac{-2136}{\sqrt{4795200}}$$

$$=\frac{-2136}{2189.8}$$

$$=-.975$$

PROBLEM 4.3. *The following data give the index numbers of industrial production of a country and the number of registered unemployed persons in the same country during the years 1964—71.*

Year	*Industrial Production (Index Number)*	*Number of registered Unemployed (Hundred thousand)*
1964	*100*	*11.3*
1965	*102*	*12.4*
1966	*104*	*14.0*
1967	*107*	*11.1*
1968	*105*	*12.3*
1969	*112*	*12.2*
1970	*103*	*19.1*
1971	*94*	*26.4*

Calculate the co-efficient of correlation between production and the number of unemployed.

SOLUTION. Calculation of the co-efficient of correlation between production and the number of unemployed.

Industrial Production				Number of registered unemployed			
Year	Index number m_1	Deviation from as. av. (105) x	Square of deviation x^2	Number (Hundred thousand) m_2	Deviation from as. av. (14.0) y	Square of deviation y^2	Product of deviation xy
1964	100	— 5	25	11.3	— 2.7	7.29	+ 13.5
1965	102	— 3	9	12.4	— 1.6	2.56	+ 4.8
1966	104	— 1	1	14.0	0	0	0
1967	107	+ 2	4	11.1	— 2.9	8.41	— 5.8
1968	105	0	0	12.3	— 1.7	2.89	0
1969	112	+ 7	49	12.2	— 1.8	3.24	— 12.6
1970	103	— 2	4	19.1	+ 5.1	26.01	— 10.2
1971	94	—11	121	26.4	+12.4	153.76	—136.4
$n=8$		$\Sigma x = -13$	$\Sigma x^2 =$ 213		$\Sigma y = +6.8$	$\Sigma y^2 =$ 204.16	$\Sigma xy =$ —146.7

Applying to the above data, the Karl Pearson's formula,

$$r = \frac{\Sigma xy - n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} \times \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)}^{2}}$$

We get,

$$= \frac{-146.7 - 8\left(\frac{-13}{8}\right)\left(\frac{6.8}{8}\right)}{8\sqrt{\frac{213}{8} - \left(\frac{-13}{8}\right)^2} \times \sqrt{\frac{204.16}{8} - \left(\frac{6.8}{8}\right)}^{2}}$$

$$= \frac{-146.7 + 11.05}{8\sqrt{23.98 \times 24.8}}$$

$$= \frac{-135.65}{8 \times 24.4}$$

$$\frac{-135.65}{195.2}$$

$$= -.695$$

Thus, the co-efficient of correlation between production and the number of unemployed is —.695.

PROBLEM. 4.4. *The following table gives the value of exports of raw cotton from India and the value of the imports of manufactured cotton goods into India during the years 1953—54 to 1971—72 :—*

(In crores of rupees)

Year	*Exports of Raw Cotton*	*Imports of manufactured cotton goods*
1953—54	*42*	*56*
1957—58	*44*	*49*
1959—60	*58*	*53*
1961—62	*55*	*58*
1963—64	*89*	*65*
1969—70	*98*	*76*
1971—72	*66*	*58*

Calculate the co-efficient of correlation between the value of the exports of raw cotton and the value of the imports of cotton manufactured goods.

SOLUTION. Calculation of the co-efficient of correlation between the value of the exports of raw cotton and the value of the imports of cotton manufactured goods.

Year	Export of Raw Cotton			Import of manufactured cotton goods			Product of deviations
	Value in crores of rupees (m_1)	Deviation from as. av. (58) (x)	Square of deviation (x^2)	Value in crores of rupees (m_2)	Deviation from as. av. (58) (y)	Square of deviation (y^2)	(xy)
1953—54	42	—16	256	56	— 2	4	32
1957—58	44	—14	196	49	— 9	81	126
1959—60	58	0	0	53	— 5	25	0
1961—62	55	— 3	9	58	0	0	0
1963—64	89	+31	961	65	+ 7	49	217
1969—70	98	+40	1600	76	+18	324	720
1971—72	66	+ 8	64	58	0	0	0
$n=7$		$\Sigma x=+46$	$\Sigma x^2=$ 3086		$\Sigma y=+9$	$\Sigma y^2=$ 483	$\Sigma xy=$ 1095

Applying to the above data, the Karl Pearson's formula—

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

We get,
$$r=\frac{1095-7\left(\frac{46}{7}\right)\left(\frac{9}{7}\right)}{7\sqrt{\frac{3086}{7}-\left(\frac{46}{7}\right)^2}\sqrt{\frac{483}{7}-\left(\frac{9}{7}\right)^2}}$$

$$=\frac{1095-59.14}{7\sqrt{397.70\times 67.35}}$$

$$=\frac{1035.86}{7\times 163.66}$$

$$=\frac{1035.86}{1145.62}$$

$$=+.904$$

Thus, the co-efficient of correlation between the value of the exports of raw cotton and the value of the imports of cotton manufactured goods is +.904.

PROBLEM 4.5. *Calculate the co-efficient of correlation between cotton cloth manufactures and woollen manufactures from the following data for July 1971 to June 1972.*

Months 1971—72	*Index of Industrial Production Cotton Manufactures*	*(Base 1966=100) Woollen Manufactures*
July	*105*	*64*
August	*108*	*63*
September	*106*	*63*
October	*100*	*62*
November	*104*	*69*
December	*105*	*61*
January	*104*	*57*
February	*111*	*55*
March	*104*	*47*
April	*114*	*45*
May	*118*	*45*
June	*117*	*50*

SOLUTION. Calculation of the co-efficient of correlation between cotton manufactures and woollen manufactures.

Months	Cotton Manu.	Dev. from as. av. (100)	Dev. Sq.	Woollen Manu.	Dev. from as. av. (50)	Dev. Sq.	Product of x and y
1971—72	(m_1)	(x)	(x^2)	(m_2)	(y)	(y^2)	(xy)
July	105	+ 5	25	64	+14	196	+ 70
August	108	+ 8	64	63	+13	169	+104
September	106	+ 6	36	66	+16	256	+ 96
October	100	0	0	62	+12	144	0
November	104	+ 4	16	69	+19	361	+ 76
December	105	+ 5	25	61	+11	111	+ 55
January	104	+ 4	16	57	+ 7	49	+ 28
February	111	+11	121	55	+ 5	25	+ 55
March	104	+ 4	16	47	— 3	9	— 12
April	114	+14	196	45	— 5	25	— 70
May	118	+18	324	45	— 5	25	— 90
June	117	+17	289	50	0	0	0
$n=12$		$\Sigma x=+96$	$\Sigma x^2=1128$		$\Sigma y=+84$	$\Sigma y^2=1380$	$\Sigma xy=312$

Co-efficient of correlation or

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma x}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

Multiplying both numerator and denominator by '*n*' we get

$$=\frac{\Sigma xy\times n-(\Sigma x)(\Sigma y)}{\sqrt{\Sigma x^2\times n-(\Sigma x)^2}\sqrt{\Sigma y^2\times n-(\Sigma y)^2}}$$

$$=\frac{312\times 12-(96\times 84)}{\sqrt{1128\times 12-(96)^2}\quad\sqrt{1380\times 12-(84)^2}}$$

$$=\frac{3744-8064}{\sqrt{4320\times 9504}}$$

$$=\frac{-4320}{\sqrt{4320\times 9504}}$$

$$=\frac{-4320}{65.73\times 97.49}\quad=\frac{-4320}{6407.04}$$

$$=-.67$$

PROBLEM 4.6. *Calculate the co-efficient of correlation between the values of x and y given below :—*

x	*y*
78	*125*
89	*137*
97	*156*
69	*112*
59	*107*
79	*136*
68	*123*
61	*108*

(You may use 69 as working mean for x and 112 as that for y)

SOLUTION. Calculation of co-efficient of correlation between x and y.

x (m_1)	Dev. from as. av. (69) (x)	(x^2)	y (m_2)	Dev. from as. av. (112) (y)	(y^2)	xy
78	+ 9	81	125	+ 13	169	+ 117
89	+20	400	137	+ 25	625	+ 500
97	+28	784	156	+ 44	1936	+1232
69	0	0	112	0	0	0
59	−10	100	107	− 5	25	+ 50
79	+10	100	136	+ 24	576	+ 240
68	− 1	1	123	+11	121	− 11
61	− 8	64	108	− 4	16	+ 32
$n=8$	$\Sigma x=+48$	$\Sigma x^2=$ 1530	$n=8$	$\Sigma y=+108$	$\Sigma y^2=+3468$	$\Sigma xy=2160$

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

$$=\frac{2160-8\left(\frac{48}{8}\right)\left(\frac{108}{8}\right)}{8\sqrt{\frac{1530}{8}-\left(\frac{48}{8}\right)^2}\sqrt{\frac{3468}{8}-\left(\frac{108}{8}\right)^2}}$$

Multiplying both the numerator and denominator by n (8), we get,

$$=\frac{2160\times8-(48\times108)}{\sqrt{1530\times8-(48)^2}\sqrt{3468\times8-(108)^2}}$$

$$=\frac{17280-5184}{\sqrt{9936\times16080}}$$

$$=\frac{12096}{\sqrt{9936\times16080}}$$

$=$Anti log [log 12096$-\frac{1}{2}$ (log 9936$+$log 16080)]

$=$Anti log [4.0828$-\frac{1}{2}$ (3.9974$+$4.2068)]

$=$Anti log [4.0828$-\frac{1}{2}$ (8.2042)]

$=$Anti log (4.0828 $-$4.1021)

$=$Anti log $\bar{1}$.9807

$=+$.9566

PROBLEM 4.7.. *Calculate the co-efficient of correlation between infant mortality and overcrowding from the following data—*

Infant mortality	*109*	*122*	*96*	*142*	*151*	*124*	*125*	*102*	*109*	*156*	*122*
Percentage of population overcrowded	*14.9*	*6.3*	*5.8*	*12.2*	*3.32*	*13.3*	*14.6*	*8.8*	*4.9*	*39.8*	*6.3*

SOLUTION. Calculation of the co-efficient of correlation between infant mortality and overcrowding.

	Infant mortality			Overcrowding		
Infant Mortality Rate (m_1)	Deviation from as. av. (122) (x)	Square of Deviation (x^2)	Percentage of population overcrowded (m_2)	Deviation from as. av. (10) (y)	Square of Deviation (y^2)	Product of Deviations (xy)
109	—13	169	14.9	+ 4.9	24.01	— 63.7
122	0	0	6.3	— 3.7	13.69	0
96	—26	676	5.8	+ 4.2	17.64	+ 109.2
142	+20	400	12·2	+ 2.2	4.84	+ 44.0
151	+29	841	33.2	+23.2	538.24	+ 672.8
124	+ 2	4	13.3	+ 3.3	10.89	+ 6.6
125	+ 3	9	14.6	+ 4.6	21.16	+ 13.8
102	—20	400	8.8	— 1.2	1.44	+ 24.0
109	—13	169	4.9	— 5.1	26.01	+ 66.3
156	+34	1156	39.8	—29.8	889.04	+1013.2
122	0	0	6.3	—3.7	13.69	0
$n=11$	$\Sigma x=+16$	$\Sigma x^2=$ 3824		$\Sigma y=+$ 50.1	$\Sigma y^2=$ 1559.65	$\Sigma xy=+$ 1886.2

Applying to the above data, the Karl Pearson's formula,

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

we get,

$$r=\frac{1886.2-11\left(\frac{16}{11}\right)\left(\frac{50.1}{11}\right)}{11\sqrt{\frac{3824}{11}-\left(\frac{16}{11}\right)^2}\sqrt{\frac{1559.65}{11}-\left(\frac{50.1}{11}\right)^2}}$$

$$=\frac{1886.2\times 11-16\times 50.1}{\sqrt{3824\times 11-(16)^2}\sqrt{1559.65\times 11-(50.1)^2}}$$

$$=\frac{20748.2-801.6}{\sqrt{42064-256}\sqrt{17156.15-2510.01}}$$

$$=\frac{19946.6}{\sqrt{41808\times 14646.14}}=\frac{19946.6}{24745.2}=+.806$$

Or=Anti-log [log 19946.6—½ (log 41808+log 14646.14)]
=Anti-log [4.2989—½ (4.6212+4.1644)]

$$=\text{Anti-log }[4.2989-\tfrac{1}{2}(8.7856)]$$
$$=\text{Anti-log }[4.2989-4.3928)]$$
$$=\text{Anti-log }\bar{1}.9061$$
$$=+.8056$$

Thus the co-efficient of correlation between infant mortality and overcrowdin is +.8056.

PROBLEM 48. *Calculate the co-efficient of correlation from the followin table, and interpret it.*

Year	*Av. daily No. of labourers (in thousands)*	*Lakhs of bales consumed by mills*
1965	*368*	*22*
1966	*384*	*21*
1967	*385*	*24*
1968	*361*	*20*
1969	*347*	*22*
1970	*384*	*26*
1971	*395*	*26*
1972	*403*	*29*
1973	*400*	*28*
1974	*385*	*27*

SOLUTION. Calculation of the co-efficient of correlation between the number of labourers and the bales of raw material consumed by the mills.

Year	Av. daily No. of labourers			Bales consumed by mills			
	No. of labourers in thousands m_1	Deviation from the as. av. (380) (x)	Square of deviation (x^2)	Bales in lakhs m_2	Deviation from the as. av. (25) (y)	Square of deviation (y_2)	Product of deviation (xy)
1965	368	−12	144	22	−3	9	+36
1966	384	+4	16	21	−4	16	−16
1967	385	+5	25	24	−1	1	−5
1968	361	−19	361	20	−5	25	+95
1969	347	−33	1089	22	−3	9	+99
1970	384	+4	16	26	+1	1	+4
1971	395	+15	225	26	+1	1	+15
1972	403	+23	529	29	+4	16	+92
1973	400	+20	400	28	+3	9	+60
1974	385	+5	25	27	+2	4	+10
$n=10$		$\Sigma x=+12$	$\Sigma x^2=2830$		$\Sigma y=-5$	$\Sigma y^2=91$	$\Sigma xy=+390$

Applying to the above data, the Karl Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x_2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

we get,

$$r=\frac{390-10\left(\frac{12}{10}\right)\left(\frac{-5}{10}\right)}{10\sqrt{\frac{2850}{10}-\left(\frac{12}{10}\right)^2}\sqrt{\frac{91}{10}-\left(\frac{-5}{10}\right)^2}}$$

$$=\frac{390\times10-(12\times-5)}{\sqrt{2830\times10-(12)^2}\ \sqrt{91\times10-(-5)^2}}$$

$$=\frac{3900+60}{\sqrt{28300-144}\ \sqrt{910-25}}$$

$$=\frac{3960}{\sqrt{28156\times885}}$$

$=\text{Anti-log}\ [(\log 3960-\frac{1}{2}(\log 28156+\log 885)]$

$=\text{Anti-log}\ [(3.5977-\frac{1}{2}(4.4502+2.9469)]$

$=\text{Anti-log}\ [(3.5977-\frac{1}{2}(7.3971)]$

$=\text{Anti-log}\ [(3.5977-3.6985)]$

$=\text{Anti-log}\ 1.8992$

$=+.7929$

There is a high degree of positive correlation between the number of labourers employed and the number of bales consumed which means that with an increase in the number of persons employed the consumption of bales increases and vice versa.

PROBLEM 4.9. *Calculate r from the following table, and indicate its probable error.*

	Net area sown in lakhs of acres	*No. of ploughs in lakhs*
U. P.	*359*	*52*
Tamil Nadu	*310*	*44*
Maharashtra	*285*	*12*
Punjab	*275*	*24*
Andhra	*257*	*35*
M. P.	*245*	*16*
W. Bengal	*240*	*46*
Assam	*64*	*11*
Bihar	*48*	*3*
Orissa	*23*	*2*
	Average 211	*25*

SOLUTION. Calculation of the co-efficient of correlation between the net area sown and the number of ploughs used in different states of India.

	Net area sown			No. of ploughs			
Provinces	Area in lakhs of acres m_1	Deviation from the as. av. (211) (x)	Square of deviation (x^2)	No. of ploughs in lakhs m_2	Deviation from the as. av. (25) (y)	Square of deviation (y^2)	Product of deviations (xy)
U. P.	359	+148	21904	52	+27	729	+3996
Tamil Nadu	310	+ 99	9801	44	+19	361	+1881
Maharashtra	285	+ 74	5476	12	—13	169	— 962
Punjab	275	+ 64	4096	24	+ 1	1	— 64
Andhra	257	+46	2116	35	+10	100	+ 460
M. P.	245	+ 34	1156	16	— 9	81	— 306
W. Bengal	240	+ 29	841	46	+21	441	+ 609
Assam	64	—147	21609	11	—14	196	+2058
Bihar	48	—163	26569	3	—22	484	+3586
Orissa	23	—188	35344	2	—23	529	+4324
$n=10$		$\Sigma x =$ —4	$\Sigma x^2=$ 128912		$\Sigma y=-5$	$\Sigma y^2=$ 3091	$\Sigma xy=$ +15582

Substituting the above values in the Karl Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

We get $r=\dfrac{+15582-10\left(\frac{-4}{10}\right)\left(\frac{-5}{10}\right)}{10\sqrt{\frac{128912}{10}-\left(\frac{-4}{10}\right)^2}\sqrt{\frac{3091}{10}-\left(\frac{-5}{10}\right)^2}}$

$$=\frac{15582\times10-(-4\times-5)}{\sqrt{128912\times10-(-4)^2}\sqrt{3091\times10-(-5)^2}}$$

$$\frac{155820-20}{\sqrt{1289120-16}\sqrt{30910-25}}$$

$$=\frac{155800}{\sqrt{1289104\times30885}}$$

=Anti log [log. 155800—½ (log.1289104+log. 30885)

=Anti log [5.1931—½ (6.1106+4.4900)]

=Anti log [5.1931—½ (10.6006)]

=Anti log [5.1931—5.3003)]

=Anti log $\bar{1}.8928$

=+.78

Probable error of the co-efficient of correlation

$$\text{P. E.} = .6745 \frac{1-r^2}{\sqrt{n}}$$

$$= .6745 \frac{1-(.78)^2}{\sqrt{10}}$$

$$= .6745 \frac{1-.6084}{3.162}$$

$$\fallingdotseq .0835$$

PROBLEM 4.10. *The following table gives the results of the Matriculation Examination held in 1972 :—*

Age of Candidetes	*13—*	*14—*	*15—*	*16—*	*17—*	*18—*	*19—*	*20—*	*21—*
Percentage of failures	*39.2*	*40.6*	*43.4*	*34.2*	*36.6*	*39.2*	*48.9*	*47.1*	*54.5*

Calculate the co-efficient of correlalion and estimate its probable error., From your results can you definitely assert that failure is correlated with age ?

SOLUTION. Calculation of the co-efficient of correlation between the age of the candidates for the Matriculation examination and the percentage of failures.

	Age of candidates			Percentage of failures		
Age in years m_1	Deviation from the as. av. (17.5) (x)	Square of deviation (x^2)	Percentage m_2	Deviation from the as. av. (40) (y)	Square of deviation (y^2)	Product of deviation (xy)
Mid-Points						
3— 13.5	—4	16	39.2	— .8	.64	+ 3.2
14— 14.5	—3	9	40.6	+ .6	.36	— 1.8
15— 15.5	—2	4	43.4	+ 3.4	11.56	— 6.8
16— 16.5	—1	1	34.2	— 5.8	33.64	+ 5.8
17— 17.5	0	0	36.6	—3.4	11.26	0
18— 18.5	+1	1	39.2	— .8	.64	— ·8
19— 19.5	+2	4	48.9	+ 8,9	79.21	+17.8
20— 20.5	+3	9	47.1	+ 7.1	50.41	+21.3
21— 21.5	+4	16	54.5	+ 14.5	210.25	+58.0
=9	$\Sigma x=0$	$\Sigma x^2=60$		$\Sigma y=$ +23.7	$\Sigma y^2=$ 398.27	$\Sigma xy=$ +96.70

Substituting the above values in the Karl Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

we get,

$$r=\frac{96.7-9\left(\frac{0}{9}\right)\left(\frac{23.7}{9}\right)}{9\sqrt{\frac{60}{9}-\left(\frac{0}{9}\right)^2}\sqrt{\frac{398.27}{9}-\left(\frac{23.7}{9}\right)^2}}$$

$$=\frac{96.7\times9-(0\times23.7)}{\sqrt{60\times9-(0)^2}\;\sqrt{398.27\times6-(23.7)^2}}$$

$$=\frac{870.3}{\sqrt{540}\;\sqrt{3584.43-561.69}}$$

$$=\frac{870\ 3}{\sqrt{540\times3022.74}}$$

$=$Anti log $[\log 870.3-\frac{1}{2}(\log 540+\log 3022.74]$

$=$Anti log $[2.9395-\frac{1}{2}(2.7324+3.4800)]$

$=$Anti log $[2.9395-\frac{1}{2}(6.2124)]$

$=$Anti log $[2.9335-3.1012]$

$=$Anti log $\bar{1}.8323$

$=+.6797$ or $+.680$

Probable Error of the co-efficient of correlation

$$\text{P.E.}=6745\left(\frac{1-r^2}{\sqrt{n}}\right)=.6745\ \frac{1-(.680)^2}{\sqrt{9}}$$

$$=\ .12$$

Thus, the co-efficient of correlation between the age of candidates and th percentage of failures is +.680 and its probable error is .12.

Co-efficient of correlation is moderately high but the probable error is als significant. Since the co-efficient of correlation is less than 6 times probable erro it cannot be definitely asserted that failure is correlated with age.

PROBLEM. 4.11. *The index number of prices of all commodities in Bomba and in Calcutta were as under :—*

Month		*Index number of commodity price in Calcutta*	*Index number of commodity prices in Bombay*
May	*1972*	*169*	*204*
June	*1972*	*182*	*222*
July	*1972*	*182*	*225*

August	*1972*	*192*	*228*
September	*1972*	*198*	*229*
ctober	*1972*	*209*	*233*
v.	*1972*	*227*	*249*
Dec.	*1972*	*238*	*266*
anuary	*1973*	*250*	*255*
eb.	*1973*	*253*	*255*

Do you think prices in Bombay and in Calcutta are correlated ?

SOLUTION. Calculation of the co-efficient of correlation between Calcutta nd Bombay Index number of commodity prices.

Month	Index number of commodity prices in Calcutta			Index number of commodity prices in Bombay			
	Index number m_1	Deviation from the as. av. (209) (x)	Square of deviation (x^2)	Index number m_2	Deviation from the as. . av. (229) (y)	Square of deviation (y^2)	Product of deviation (xy)
May 1972	169	−40	1600	204	—25	625	1000
June 1972	182	—27	729	222	— 7	47	189
July 1972	182	—27	729	225	— 4	16	108
Aug. 1972	192	—17	289	228	— 1	1	17
Sep. 1972	198	—11	121	229	0	0	0
Oct. 1972	209	0	0	233	+ 4	16	0
Nov. 1972	227	+18	324	249	+20	400	360
Dec. 1972	238	+29	841	266	+37	1369	1073
Jan. 1973	250	+41	1681	255	+26	676	1066
Feb. 1973	253	+44	1936	255	+26	676	1144
$n-10$		$\Sigma x=$ +10	$\Sigma x^2=$ 8250		$\Sigma y=$ +76	$\Sigma y^2=$ 3828	$\Sigma xy=$ 4957

Applying to the above data the Karl Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma y}{n}\right)}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

$$r=\frac{4957-10\left(\frac{10}{10}\right)\left(\frac{76}{10}\right)}{10\sqrt{\frac{8250}{10}-\left(\frac{10}{10}\right)^2}\sqrt{\frac{3828}{10}-\left(\frac{76}{10}\right)^2}}$$

$$=\frac{4957-76}{10\sqrt{825-1}\sqrt{382.8-57.8}}$$

$$=\frac{4881}{10\sqrt{824\times 325}}$$

$$=\frac{4881}{10\sqrt{824\times 325}}$$

$=.944$

$=\text{Anti log }[\log 4881-\tfrac{1}{2}(\log 824+\log 325)]$

$=\text{Anti log }[2.6884-\tfrac{1}{2}(2.9159+2.5119)]$

$=\text{Anti log }[2.6884-2.7139]$

$=\text{Anti log }\bar{1}.9745$

$= .9441$ or $.944$

PROBLEM 4.12. *The following table gives the distribution of the total p lation and those who are wholly or partially blind among them. Find out if th any relation between age and blindness :—*

Age	*No. of persons in thousands*	*Blind*
0—10	*100*	*55*
10—20	*60*	*67*
20—30	*40*	*100*
30—40	*36*	*111*
40—50	*24*	*150*
50—60	*11*	*200*
60—70	*6*	*300*
70—80	*3*	*500*

SOLUTION. Calculation of the co-efficient of correlation between ag blindness.

	Age			Blindness			
Years Age group	Mid-Points m_1	Deviation from the as. as. (45) (x)	Square of Deviation (x^2)	Blind per lakh m_2	Deviation from the as. av. (185 (y)	Square of Deviation (y^2)	Product deviatio (x
0—10	5	—40	1600	55	—130	16900	+52
10—20	15	—30	900	67	—118	13924	+35
20—30	25	—20	400	100	— 85	7225	+17
30—40	35	—10	100	111	— 74	5476	+ 7
40—50	45	0	0	150	— 35	1225	
50—60	55	+10	100	200	+ 15	225	+ 1
60—70	65	+20	400	300	+115	13225	+23
70—80	75	+30	900	500	+315	99225	+94
$n = 8$		$\Sigma x=$ −40	$\Sigma x^2=$ 4400		$\Sigma y=+3$	$\Sigma y^2=$ 157425	$\Sigma xy=$ +2308

Applying to the above data, the Karl Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\;\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

We get,

$$r=\frac{23080-8\left(\frac{-40}{8}\right)\left(\frac{+3}{8}\right)}{\sqrt{\frac{4400}{8}-\left(\frac{-40}{8}\right)}^{2}\;\sqrt{\frac{157425}{8}-\left(\frac{3}{8}\right)}^{2}}$$

$$r=\frac{23080+15}{\sqrt{550-25}\;\sqrt{157425\times 8-(3)^2}}$$

$$r=\frac{23095}{\sqrt{525}\;\sqrt{1259400-9}}$$

$$=\frac{23095}{\sqrt{525\times 1259391}}$$

$=$Anti log [log 23095 $-\frac{1}{2}$ (log 525 + log 1259391)]

$=$Anti log [4.3636 $-\frac{1}{2}$ (2.7202 + 6.1004)]

$=$Anti log [4.3636 $-$ 4.4103]

$=$Anti log $\bar{1}.9533$

$= + .898.$

There is a definite relationship between age and blindness. Blindness increases with increase in age.

PROBLEM 4.13. *The following marks were obtained by 12 students in Mathematics and Statistics,*

Students	*Mathematics*	*Statistics*
A	*50*	*22*
B	*54*	*25*
C	*56*	*34*
D	*59*	*28*
E	*60*	*26*
F	*62*	*30*
G	*61*	*32*
H	*65*	*30*
I	*67*	*28*
J	*71*	*34*
K	*71*	*36*
L	*74*	*40*

Find the co-efficient of correlation between the performance of the students in Mathematics and Statistics, as determined by these data.

SOLUTION. Calculation of the co-efficient of correlation between the perfor-mance of students in Mathematics and Statistics.

Students	Marks in Mathematics			Marks in Statistics			
	Marks of each student in Math. m_1	Deviations from Assumed Average (62) (x)	Square of Deviations (x^2)	Marks of each student in Statistics m_2	Deviations from Assumed Average (30) (y)	Square of Deviations (y^2)	Product of Deviations (xy)
A	50	−12	144	22	− 8	64	+ 9
B	54	− 8	64	25	− 5	25	+ 4
C	56	− 6	36	34	+ 4	16	− 2
D	59	− 3	9	28	− 2	4	+
E	60	− 2	4	26	− 4	16	+
F	62	0	0	30	0	0	
G	61	− 1	1	32	+ 2	4	−
H	65	+ 3	9	30	0	0	
I	67	+ 5	25	28	− 2	4	− 1
J	71	+ 9	81	34	+ 4	16	+ 3
K	71	+ 9	81	36	+ 6	36	+ 5
L	74	+12	144	40	+10	100	+12
$n=12$		$\Sigma x=+6$	$\Sigma x^2=$ 598		$\Sigma y=+5$	$\Sigma y^2=$ 285	$\Sigma xy=$ +324

Applying to the above data, the Karl Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

$$r=\frac{+324-12\left(\frac{6}{12}\right)\left(\frac{5}{12}\right)}{12\sqrt{\frac{598}{12}-\left(\frac{6}{12}\right)^2}\sqrt{\frac{285}{12}-\left(\frac{5}{8}\right)^2}}$$

$$r=\frac{+324\times12-(6\times5)}{\sqrt{598\times12-(6)^2}\sqrt{285\times12-(5)^2}}$$

$$r=\frac{3888-30}{\sqrt{7176-36}\sqrt{3420-25}}$$

$$r=\frac{3858}{\sqrt{7140\times3395}}$$

$r=$Anti log [log 3858$-\frac{1}{2}$ (log 7140+log 3395)]

$r=$Anti log [3.5864$-\frac{1}{2}$ (3.8537+3.5308)]

$r = \text{Anti log}\ [3.5864 - \tfrac{1}{2}\ (7.3845)]$

$r = \text{Anti log}\ [3.5864 - 3.6922]$

$r = \text{Anti log}\ \bar{1}.8942$

$r = +\ .783.$

PROBLEM 4.14 *Find the co-efficient of correlation from the following table :*

x	*5*	*10*	*15*	*20*	*25*	*30*	*Total*
10	...	*1*	*1*	*2*	*8*	*12*	*24*
15	*1*	*2*	*5*	*9*	*80*	*11*	*108*
20	*2*	*15*	*42*	*98*	*36*	*8*	*201*
25	*5*	*20*	*51*	*37*	*10*	*2*	*125*
30	*8*	*16*	*8*	*5*	*4*	*1*	*42*
Total	*6*	*54*	*107*	*151*	*138*	*34*	*500*

SOLUTION. Calculation of the co-efficient of correlation between *x* and *y*.

						x						
			5	10	15	20	25	30				
		dx	−15	−10	−5	0	+5	+10	Total			
	dy		−3	−2	−1	0	+1	+2	*f*	*fdy*	fdy^2	*fdxdy*
10	—10	—2	0 0	4 1	2 1	0 2	−16 8	−48 12	24	—48	96	—58
15	— 5	—1	3 1	4 2	5 5	0 9	−80 80	−22 11	108	-108	108	—90
20	0	0	0 2	0 15	0 42	0 98	0 36	0 8	201	0	0	0
25	+ 5	+1	−15 5	−40 20	51 51	0 37	10 10	4 2	125	125	125	—92
30	+10	+2	−48 8	−64 16	−16 8	0 5	8 4	4 1	42	84 Σfdy	168 Σfdy^2	—116
Total *f*			16	54	107	51	138	34	500 *fdx*	=53	=497	
fdx			-48	-108	-107	0	138	68	=-57 Σfdx^2			
fdx^2			44	216	107	0	138	136	=741			
fdxdy			−60	—96	—60	0	—78	—62		$\Sigma fdxdy = —356$		

Applying to the above data, Karl Pearson's formula

$$r = \frac{\Sigma fdxdy - n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdy}{n} - \left(\frac{\Sigma fdy}{n}\right)^2}}$$

$$= \frac{-356 - 500\left(\frac{-57}{500}\right)\left(\frac{53}{500}\right)}{500\sqrt{\frac{741}{500} - \left(\frac{-57}{500}\right)^2}\ \sqrt{\frac{497}{500} - \left(\frac{53}{500}\right)^2}}$$

$$= \frac{-356 \times 500 - (-57 \times 53)}{\sqrt{741 \times 500 - (-57)^2}\ \sqrt{497 \times 500 - (53)^2}}$$

$$= \frac{-178000 + 302'}{\sqrt{370500 - 3249}\ \sqrt{248500 - 2809}}$$

$$= \frac{174979}{\sqrt{367251 - 245691}}$$

$= \text{Anti-log}\ [\log 174979 - \frac{1}{2}(\log 367251 + \log 245691)]$

$= \text{Anti-log}\ [5{,}2430 - \frac{1}{2}(5.5649 + 5.3909)]$

$= \text{Anti-log}\ [5.2430 - 5.4779]$

$= \text{Anti-log}\ \bar{1}.765$

$= +.5822$

Thus, the co-efficient of correlation between X and Y is $+ .5822$.

PROBLEM 4.15. *Calculate the co-efficient of correlation between the ages of 100 mothers and daughters from the following data :*

Age of mother in years	*Age of daughters in years*					*Total*
	5—10	*10—15*	*15—20*	*20—25*	*25—30*	
15—25	*6*	*3*	...	...	...	*9*
25—35	*3*	*16*	*10*	...	...	*29*
35—45	...	*10*	*15*	*7*	...	*32*
45—55	...	...	*7*	*10*	*4*	*21*
55—65	...	...	...	*4*	*5*	*9*
Total	*9*	*29*	*32*	*21*	*9*	*100*

SOLUTION. Calculation of the co-efficient of correlation between the ages of mothers and daughters.

Age of mothers in years				Age of daughters in years								
				Age group 5—10	10-15	15-20	20-25	25-30				
				Mid-points 7.5	12.5	17.5	22.5	27.5				
			dx	—10	—5	0	+ 5	+10	Total			
dy				— 2	—1	0	+ 1	+ 2	f	fdy	fdy^2	fdxdy
Age group 15-25	M.P 20	--20	—2	24 6	6 3				9	—18	36	30
25-35	30	—10	—1	6 3	16 16	0 10			29	—29	29	22
35-45	40	0	0	0	0 10	0 15	0 7		32	0	0	0
45-55	50	+10	+2			0 7	10 10	8 4	21	+21	21	18
55-65	60	+20	+2				8 4	20 5	9	+18	36	28
Total f				9	29	32	21	9	100	Σfdy =—8	Σfdy^2 =122	
fdx				18	29	0	+21	+18	Σfdx =—8			
fdx^2				36	29	0	21	36	Σfdx^2 =122			
fdxdy				30	22	0·	18	28			fdxdy=	98

Substituting the above values in the Pearsen's formula

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdy^2}{n}-\left(\frac{\Sigma fdy}{n}\right)}}$$

We get,

$$r=\frac{98-100\left(\frac{-8}{100}\right)\left(\frac{-8}{100}\right)}{100\sqrt{\frac{122}{100}-\left(\frac{-8}{100}\right)^2}\sqrt{\frac{122}{100}-\left(\frac{-8}{100}\right)^2}}$$

$$=\frac{98\times100(-8\times-8)}{\sqrt{122\times100-(-8)^2}\sqrt{122\times100-(-8)^2}}$$

$$=\frac{9800-64}{12200-64}$$

$$=\frac{9736}{12136}$$

$$=+.802$$

Thus the co-efficient of correlation between the ages of mothers and daughters is +.802.

PROBLEM 4.16 *The following table gives the number of students having different heights and weights.*

Height in Centimeters	*Weight in kilograms*					
	55—60	*60—65*	*65—70*	*70—75*	*75—80*	
150—155	*1*	*3*	*7*	*5*	*2*	*18*
155—160	*2*	*4*	*10*	*7*	*4*	*27*
160—165	*1*	*5*	*12*	*10*	*7*	*35*
165—170	...	*3*	*8*	*6*	*3*	*20*
Total	*4*	*15*	*37*	*28*	*16*	*100*

Do you find any relation between height and weight ?

SOLUTION. Calculation of the co-efficient of correlation between the heights and weights of the students.

Height in centimetres				Weight in kilograms								
				Group 55—60	60—65	65—70	70—75	75—80				
				Mid-Points 57.5	62.5	67.5	72.5	77.5				
			dx	—10	—5	0	+5	+10	Total			fdxdy
		dy		— 2	—1	0	+1	+2	f	fdy	fdy^2	
Height Group 150—155	M. P. 152.5	— 5	—1	2 1	3 3	0 7	—5 5	—4 2	18	—18	18	-4
155—160	157.5	0	0	0 2	0 4	0 10	0 7	0 4	27	0	0	0
160—165	162.5	+5	+1	—2 1	—5 5	0 12	10 10	14 7	35	35	35	17
165—170	167.5	+10	+2	0 0	—6 3	0 8	12 6	12 3	20	40	80	18
Total f				4	15	37	28	16	100	Σfdy = 57	Σfdy^2 = 133	
fdx				—8	-15	0	28	32	Σfdx = 37			
fdx^2				16	15	0	28	64	Σfdx^2 = 123			
fdxdy				0	—8	0	17	22		Σfdxdy = 31		

Substituting the above values in the Pearson's formula

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdy^2}{n}-\left(\frac{\Sigma fdy}{n}\right)^2}}$$

We get,

$$=\frac{31-100\left(\frac{37}{100}\right)\left(\frac{57}{100}\right)}{100\sqrt{\frac{123}{100}-\left(\frac{37}{100}\right)^2}\sqrt{\frac{133}{100}-\left(\frac{57}{100}\right)^2}}$$

$$r = \frac{31 \times 100 - 37 \times 57}{\sqrt{123 \times 100 - (37)^2}\ \sqrt{133 \times 100 - (57)^2}}$$

$$= \frac{991}{\sqrt{10931 \times 10051}}$$

$$= \text{Anti log}\left[\log 991 - \tfrac{1}{2}(\log 10931 + \log 10051)\right]$$

$$= +.0945.$$

The co-efficient of correlation between height and weight is +.0945. It indicates that there is no relationship between the two phenomena.

PROBLEM 4.17 *The following table gives the frequency according to age groups of marks obtained by 67 students in an intelligence test :—*

	Age in years				
Test marks	*18*	*19*	*20*	*21*	*Total*
200—250	*4*	*4*	*2*	*1*	*11*
250—300	*3*	*5*	*4*	*2*	*14*
300—350	*2*	*6*	*8*	*5*	*21*
350—400	*1*	*4*	*6*	*10*	*21*
Total	*10*	*19*	*20*	*18*	*67*

Is there any relationship between age and intelligence ?

SOLUTION. Calculation of the co-efficient of correlation between age and intelligence.

Test marks				Age in years							
				18	19	20	21				
			dx	—2	—1	0	+1				
			dy					Total f	fdy	fdy^2	$fdxdy$
Groups	M.P.										
200—250	225	—100	—2	16 4	8 4	0 2	—2 1	11	—22	44	22
250—300	275	—50	—1	6 3	5 5	0 4	—2 2	14	—14	14	9
300—350	325	0	0	0 2	0 6	0 8	0 5	21	0	0	0
350—400	375	+50	+1	—2 1	—4 4	0 6	10 10	21	+21	21	4
Total f				10	19	20	18	67	$\Sigma fdy = -15$	$\Sigma fdy^2 = 79$	
fdx				—20	—19	0	+18	$\Sigma fdx = --21$			
fdx^2				40	19	0	18	$\Sigma fdx^2 = 77$			
$fdxdy$				20	9	0	6		$\Sigma fdxdy = 35$		

Substituting the above values in the Pearson's formula.

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdy^2}{n}-\left(\frac{\Sigma fd}{n}\right)}}$$

$$r=\frac{35-67\left(\frac{-21}{67}\right)-\left(\frac{-15}{67}\right)}{67\sqrt{\frac{77}{67}-\left(\frac{-21}{67}\right)^2}\sqrt{\frac{79}{67}-\left(\frac{-15}{67}\right)^2}}$$

$$=\frac{35\times67-(-21\times-15)}{\sqrt{77\times67-(-21)^2}\sqrt{79\times67-(-15)^2}}$$

$$=\frac{2030}{\sqrt{4718\times5068}}$$

$$=\text{Anti log }[\log.\ 2030-\tfrac{1}{2}(\log 4718+\log 5068)]$$

$$=+.415$$

The co-efficient of correlation indicates a moderate degree of relationship between age and intelligence. Since the co-efficient of correlation is positive it means that intelligence increases with an increase in age.

PROBLEM 4.18 *Calculate from the data reproduced below pertaining to 66 selected villages in Meerut District, the value of r between 'total cultivable area' and 'the area under wheat.'*

Total Cultivable Area (in Acres)

Area under wheat (in Acres)	*0—500*	*500—1000*	*1000—1500*	*1500—2000*	*2000—2,500*	*Total*
0—200	*12*	*6*	..	..	..	*18*
200—400	*2*	*18*	*4*	*2*	*1*	*27*
400—600	..	*4*	*7*	*3*	..	*14*
600—800	..	*1*	..	*2*	*1*	*4*
800—1000	..	..	..	*1*	*2*	*3*
Total	*14*	*29*	*11*	*8*	*4*	*66*

SOLUTION. Calculation of co-efficient of correlation between total are and area under wheat.

Area under Wheat (in Acres)				Total Cultivable Area (in Acres)								
				Groups 0--500	500—1000	1000—1500	1500—2000	2000—2500				
				Mid-points 250	750	1250	1750	2250				
		dy		—1000	—500	0	+500	+1000	Total			
		dx		—2	—1	0	+1	+2	*f*	*fdy*	fdy^2	*fdx dy*
Groups 0—200	MP 100	—400	—2	48 12	12 6	..	..	..	18	—36	72	6
200—400	300	—200	—1	4 2	18 18	0 4	—2 2	—2 1	27	—27	27	18
400—600	500	0	0	..	0 4	0 7	0 3	..	14	0	0	0
600—800	700	+200	+1	..	—1 1	..	2 2	2 1	4	+4	4	3
800—1000	900	+400	+2	..	..	..	2 1	8 2	3	+6	12	10
Total *f*				14	29	11	8	4	66	$\Sigma fdy = -53$	$\Sigma fdy^2 = 115$	
fdx				—28	29	0	+8	+8	$\Sigma fdx = -41$			
fdx^2				56	29	0	8	16	$\Sigma fdx^2 = 109$			
fdxdy				52	- 29	0	2	8				$\Sigma fdxdy = 91$

Substituting the above values in the Pearson's formula

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)}\sqrt{\frac{\Sigma fdy^2}{n}-\left(\frac{\Sigma fdy}{n}\right)^2}}$$

$$r=\frac{91-66\left(\frac{-41}{66}\right)\left(\frac{-53}{66}\right)}{66\sqrt{\frac{109}{66}-\left(\frac{-41}{66}\right)^2}\sqrt{\frac{115}{66}-\left(\frac{-53}{66}\right)^2}}$$

$$r=\frac{91\times66-(-41\times-53)}{\sqrt{109\times66-(-41)^2}\sqrt{115\times66-(-53)^2}}$$

$$=\frac{6006-2173}{\sqrt{7194-1681}\sqrt{7590-2809}},$$

$$=+.746.$$

Thus the co-efficient of correlation between 'total cultivable area' and 'the area under wheat' is +.746.

PROBLEM 4.19 *The correlation table given below shows the ages of husband and wife for 53 married couples living together on the census night of 1971. Calculate the co-efficient of correlation between the age of husband and that of his wife.*

	Age of wife						
Age of Husband	*15—25*	*25—35*	*35—45*	*45—55*	*55—65*	*65—75*	*Total*
15—25	*1*	*1*	..	..	..	..	*2*
25—35	*2*	*12*	*1*	..	..	..	*15*
35—45	..	*4*	*10*	*1*	..	..	*15*
45—55	..	..	*3*	*6*	*1*	..	*10*
55—65	..	..	..	*2*	*4*	*2*	*8*
65—75	..	..	..	..	*1*	*2*	*3*
Total	*3*	*17*	*14*	*9*	*6*	*4*	*53*

SOLUTION. Calculation of the co-efficient of correlation between the age of husband and that of his wife.

Age of husband				Age of wife									
				Age group 15—25	25—35	35—45	45—55	55—65	65-75				
				Mid-points 20	30	40	50	60	70				
			dy	—20	—10	0	+10	+20	+30	Total			
dx				—2	—1	0	+1	+2	+3	*f*	*fdy*	*fdy*²	*fdxdy*
Age group 15—25	M. P. 20	—20	—2	4 1	2 1					2	— 4	8	6
25—35	30	—10	—1	4 2	12 12	0 1				15	—15	15	16
35—45	40	0	0		0 4	0 10	0 1			15	0	0	0
45—55	50	+10	+1			0 3	6 6	2 1		10	10	10	8
55—65	60	+20	+2				4 2	16 4	12 2	8	16	32	32
65—75	70	+30	+3					6 1	18 2	3	9	27	24
Total *f*				3	17	14	9	6	4	53	Σ*fdy* = 16	Σ*fdy*² =92	
fdx				—6	—17	0	9	12	12	Σ*fdx* = 10			
*fdx*²				12	17	0	9	24	36	Σ*fdx*² = 98			
fdxdy				8	14	0	10	24	30		Σ*fdxdy* = 86		

Substituting the above values in the Pearson's formula

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdy^2}{n}-\left(\frac{\Sigma fdy}{n}\right)^2}}$$

We get,

$$r=\frac{85-53\left(\frac{10}{53}\right)\left(\frac{16}{53}\right)}{53\sqrt{\left(\frac{98}{53}-\frac{10}{53}\right)}\sqrt{\left(\frac{92}{53}-\frac{66}{53}\right)}^{2}}$$

$$=\frac{86\times53-10\times16}{\sqrt{98\times52-(10)^2}\ \sqrt{92\times53-(16)^2}}$$

$$=\frac{4558-160}{\sqrt{5194-100}\ \sqrt{4876-256}}$$

$$=\frac{4398}{\sqrt{5094\times4620}}$$

$= \text{Anti log } [\log 4398-\frac{1}{2}(\log 5094+\log 4620)]$

$= \text{Anti log } [3.6434-\frac{1}{2}(3.7070+3.6646)]$

$= \text{Anti log } [3.6434-\frac{1}{2}(7.3716)]$

$= \text{Anti log } [3.6435-3.6858]$

$= \text{Anti log } \bar{1}.9577$

$= +.9072$

$= +\cdot907$

Thus the co-efficient of correlation between the age of husband and that of his wife is +.907.

PROBLEM 4.20 *Compute the co-efficient of correlation from the following table, showing the age in years of the students and the marks obtained.*

x Series→ Age in years / *y Series Marks ↓*	*16—18*	*18—20*	*20—22*	*22—24*	*Total of frequencies for x series*
10—20	2	*1*	*1*	..	*4*
20—30	*3*	*2*	*3*	*2*	*10*
30—40	*3*	*4*	*5*	*6*	*18*
40—50	*2*	*2*	*3*	*4*	*11*
50—60	..	*1*	*2*	*2*	*5*
60—70	..	*1*	*2*	*1*	*4*
Total of frequencies for y series	*10*	*11*	*16*	*15*	*52*

SOLUTION. Calculation of the co-efficient of correlation between age an intelligence.

Marks				Age in years							
				Age-group 16—18	18—20	20—22	22—24				
				Mid-points 17	19	21	23				
			dy	−4	−2	0	+2	Total			
	dx			−2	−1	0	+1	*f*	*fdy*	*fdy*^{2}	*fdxdy*
Marks group 10—20	M. P. 15	−20	−2	8 2	2 1	0 1		4	− 8	16	10
20—30	25	−10	−1	6 3	2 2	0 3	− 2 2	10	−10	10	6
30—40	35	0	0	0 3	0 4	0 5	0 6	18	0	0	0
40—50	45	+10	+1	−4 2	−2 2	0 3	4 4	11	+11	11	−2
50—60	55	+20	+2	..	—2 1	0 2	4 2	5	+10	20	2
60—70	65	+30	+3	...	—3 1	0 2	3 1	4	+12	36	0
Total *f*				10	11	16	15	52	*Σfdy* = 15	*Σfdy*^{2} = 93	
fdx				—20	—11	0	+15	*Σfdx* = —16			
fdx^2				40	11	0	15	*Σfdx*^{2} = 66			
fdxdy				10	—3	0	9			*fdxdy* = 16	

Substituting the above values in the Pearson's formula

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdy^2}{n}-\left(\frac{\Sigma fdy}{n}\right)^2}}$$

We get,

$$r=\frac{16-52\left(\frac{-16}{52}\right)\left(\frac{15}{52}\right)}{52\sqrt{\frac{66}{52}-\left(\frac{-16}{52}\right)^{2}}\sqrt{\frac{93}{52}-\left(\frac{15}{52}\right)^{2}}}$$

$$r=\frac{16\times 52-(-16\times 15)}{\sqrt{66\times 52-(-16)^{2}}\ \sqrt{93\times 52-(15)^{2}}}$$

$$=\frac{832+240}{\sqrt{3432-256}\ \sqrt{4836-225}}$$

$$=\frac{1072}{\sqrt{3176\times 4611}}$$

$= \text{Anti log }[\log 1072-\frac{1}{2}(\log 3176+\log 4611)]$

$= \text{Anti log }[3.0302-\frac{1}{2}(3.5019+3.6638)]$

$= \text{Anti log }[3.0302-\frac{1}{2}(7.1657)]$

$= \text{Anti log }(3.0302-3.5828)$

$= \text{Anti log }\bar{1}.4474$

$= +.2802$

$= +.28$

Thus the co-efficient of correlation between the age of the students and the marks obtained by them is + .28.

PROBLEM 4.21 *From the following table calculate the co-efficient of correlation between the ages of fathers and sons and estimate the probable error of the result obtained.*

Age of Fathers	*Age of Sons*								*Total*
Years	*2*	*6*	*10*	*14*	*18*	*22*	*26*	*30*	
55—60						*6*	*5*	*3*	*14*
50—55					*8*	*10*	*6*	*2*	*26*
45—50				*2*	*13*	*8*	*4*		*27*
40—45				*14*	*18*	*3*			*35*
35—40			*15*	*20*	*8*				*43*
30—35	*6*	*12*	*25*	*16*					*59*
25—30	*15*	*26*	*20*	*1*					*62*
20—25	*22*	*10*	*2*						*34*
Total	*43*	*48*	*62*	*53*	*47*	*27*	*15*	*5*	*300*

SOLUTION. Calculation of the co-efficient of correlation between ages fathers and sons.

Age of Fathers				Age of Sons										
				2	6	10	14	18	22	26	30			
dx				—16	—12	—8	—4	0	+4	+8	+12	*f* To-tal	*fdy*	*fdy*²
dy				—4	—3	—2	—1	0	+1	+2	+3			
Age group 55—60	M. P. 57.5	+20	+4						+24 6	+40 5	+36 3	14	56	224
50—55	52.5	+15	+3					0 8	+30 10	+36 6	+18 2	26	78	234
45—50	47.5	+ 1	+2				—4 2	0 13	+16 8	+16 4		27	54	108
40—45	42.5	+ 5	+1				—14 14	0 18	+3 3			35	35	35
35—40	27.5	0	0			0 15	0 20	0 8				43	0	0
30—35	32.5	— 5	—1	+ 24 6	+ 36 12	+50 25	+16 16					59	—59	59
25—30	27.5	—10	—2	+120 15	+156 26	+80 20	+2 1					62	-124	248
20—25	22.5	—15	—3	+264 22	+90 10	+12 2						34	-102	306
Total *f*				43	48	62	53	47	27	15	5	300	Σ *fdy* = —62	Σ *fdy*² = 1214
fdx				—172	—144	—124	—53	0	27	30	15	Σ *fdx* = -421		
*fdx*²				688	432	248	53	0	27	60	45	Σ *fdx*² = 1553		
fdxdy				408	282	142	0	0	73	2	54		*fdxdy* = 1	

Substituting the above values in the Pearson's formula

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdy}{n}\right)^2}}$$

where r represents the co-efficient of correlation

we get,

$$r=\frac{1051-300\left(\frac{-421}{300}\right)\left(\frac{-62}{300}\right)}{300\sqrt{\frac{1553}{300}-\left(\frac{-421}{300}\right)^2}\sqrt{\frac{1214}{300}-\left(\frac{-62}{300}\right)^2}}$$

$$=\frac{1051\times300-421\times62}{\sqrt{1553\times300-(-421)^2}\sqrt{1214\times300-(-62)^2}}$$

$$=\frac{315300-26102}{\sqrt{465900-177241}\sqrt{364200-3844}}$$

$$=\frac{289198}{\sqrt{288659\times360356}}$$

$= \text{Anti log} [\log 289198 - \tfrac{1}{2} (\log 288659 + \log 360356)]$
$= \text{Anti log} [5.4611 - \tfrac{1}{2} (5.4604 + 5.5567)]$
$= \text{Anti log} [5.4611 - \tfrac{1}{2} (11.0171)]$
$= \text{Anti log} [5.4611 - 5.5085]$
$= \text{Anti log}. \bar{1}.9526$
$= +.8966$

Probable Error of the Co-efficient of Correlation

$$\text{P. E.} = .6745\frac{1-r^2}{\sqrt{n}}$$

$$= .6745\,\frac{1-(.896)^2}{\sqrt{300}}$$

$$= .6745\,\frac{1-.80}{17.32}$$

$$= \frac{.6745\times.2}{17.32}$$

$$= .007789$$

Thus the co-efficient of correlation between the ages of father and son is +.895 and its probable error is. 007789.

PROBLEM 4.22 *Calculate r between Production of Pig Iron and Industrial Production from the following table :—*

		Pig Iron Production								
		50—60	*60—70*	*70—80*	*80—90*	*90—100*	*100—110*	*110—120*	*120—130*	*Total*
Industrial Production	*120—130*								*15*	*15*
	110—120						*6*	*34*	*1*	*41*
	100—110					*5*	*51*	*6*		*62*
	90—100				*3*	*33*	*1*			*37*
	80—90			*2*	*24*	*3*				*29*
	70—80			*7*	*2*					*9*
	60—70		*2*	*1*						*3*
	50—60	*6*	*2*							*8*
	Total	*6*	*4*	*10*	*29*	*41*	*58*	*40*	*16*	*204*

SOLUTION.

Calculation of the co-efficient of correlation between production of Pig Iron and Industrial Production

Industrial Production				Pig Iron Production (percentage of trend)											
				Group 50—60	60—70	70—80	80—90	90—100	100—110	110—120	120—130				
				Mid-points 55	65	75	85	95	105	115	125				
		dx		—40	—30	—20	—10	0	+10	+20	+30	Total			
		dy		— 4	— 3	— 2	— 1	0	+ 1	+ 2	+ 3	*f*	*fdy*	*fdy*^{2}	*fddy*
Group 120—130	M.P. 125	+30	+3								135 15	15	+45	135	135
110—120	115	+20	+2						12 6	136 34	6 1	41	+82	164	154
100—110	105	+10	+1					0 5	51 51	12 6		62	+62	62	63

90—100	95	0	0				3 ^{0}	33 ^{0}	1 ^{0}			37	0	0	0
80—90	85	−10	−1			2 ^{4}	24 ^{24}	3 ^{0}				29	−29	29	28
70—80	75	−20	−2			7 ^{28}	2 ^{4}					9	−18	36	32
60—70	65	−30	−3		2 ^{18}	1 ^{6}						3	−9	27	24
50—60	55	−40	−4	6 ^{96}	2 ^{24}							8	−32	128	120
	Total *f*			6	4	10	29	41	58	40	16	204	Σfdy = 101	Σfdy^2 = 581	
	fdx			−24	−12	−20	−29	0	+58	+80	+48	Σfdx = 101			
	fdx^2			96	36	40	29	0	58	160	144	Σfdx^2 = 563			
	fdxdy			96	42	38	28	0	63	148	141			*fdxdy* = 556	

Substituting tne above values in the Pearson's formula

$$r=\frac{\Sigma fdxdy-n\left(\frac{\Sigma fdx}{n}\right)\left(\frac{\Sigma fdy}{n}\right)}{n\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}\sqrt{\frac{\Sigma fdy^2}{n}-\left(\frac{\Sigma fdy}{n}\right)^2}}$$

We get,

$$r=\frac{556-204\left(\frac{101}{204}\right)\left(\frac{101}{204}\right)}{204\sqrt{\frac{563}{204}-\left(\frac{101}{204}\right)^2}\sqrt{\frac{581}{204}-\left(\frac{101}{204}\right)}}$$

$$=\frac{556\times 204-101\times 101}{\sqrt{563\times 204-(101)^2}\ \sqrt{581\times 204-(101)^2}}$$

$$=\frac{113424-10201}{\sqrt{114852-10201}\ \sqrt{118524-10201}}$$

$$=\frac{103223}{\sqrt{104651-108323}}$$

= Anti log [log 103223—½(log 104651 + log 108323)]
= Anti log [5.0136—½ (5.0195 + 5.0346)]
= Anti log]5.0128—½ (10.0541)]
= Anti log [5.0128—5.0271]
= Anti log $\bar{1}.9857$
= +.9677.

Thus, the co-efficient of correlation between production of Pig Iron and Industrial Production is + .9677.

PROBLEM 4.23. *Calculate the co-efficient of correlation from the following data by the method of rank differences.*

x— 75, 88, 95, 70, 60, 80, 81, 50
y—120, 134, 150, 115, 110, 140, 142, 100

SOLUTION. Calculation of co-efficient of correlation by the method of rank differences.

x	Rank	y	Rank	Rank differences d	Square of rank differences d^2
75	5	120	5	0	0
88	2	134	4	−2	4
95	1	150	1	0	0
70	6	115	6	0	0
60	7	110	7	0	0
80	4	140	3	0	0
81	3	142	2	1	1
50	8	100	8	1	1
				0	0
$n=8$				$\Sigma d=0$	$\Sigma d^2=6$

Co-efficient of correlation or $r = 1 — \frac{6(\Sigma d^2)}{n(n^2—1)} = 1 — \frac{6 \times 6}{8(64—1)}$

$= 1 — \frac{36}{504} = +.93$

Note. In this method the biggest item gets the first rank, the next biggest
cond rank and so on.

PROBLEM 4.24 *Calculate the co-efficient of rank correlation of the follo-*
ing data :—

x— 87, 22, 35, 75, 37.
y— 29, 63, 52, 46, 48.

SOLUTION. Calculation of the co-efficient of Rank Correlation.

x	Rank	*y*	Rank	Rank differences (*d*)	Square of -rank differences $(d)^2$
87	1	29	5	—4	16
22	5	63	1	+4	16
33	4	52	2	+2	4
75	2	46	4	—2	4
37	3	48	3	0	0
=5				$\Sigma d = 0$	$\Sigma d^2 = 40$

Co-efficient of correlation or $r = 1 — \frac{6(\Sigma d^2)}{n(n^2—1)}$

$= 1 — \frac{6 \times 40}{5 \times (5^2—1)}$

$= 1 — \frac{6 \times 40}{5 \times 24}$

$= 1 — \frac{240}{120}$

$= 1—2$

$= —1$

Note. This shows an absolute Negative Correlation, or perfect inverse
rrelation.

PROBLEM 4.25 *Calculate the rank co-efficient of correlation of the*
llowing data :

x— 80, 78, 75, 75, 68, 67, 60, 59.
y— 12, 13, 14, 14, 14, 16, 15, 17.

SOLUTION. Calculation of the Rank Co-efficient of Correlation.

x	Rank	*y*	Rank	Rank Differences (*d*)	Square of rank differences (d^2)
80	1	12	8	—7	49.00
78	2	13	7	—5	25.00
75	3.5	14	5	—1.5	2.25
75	3.5	14	5	—1.5	2.25
68	5	14	5	0	0.00
67	6	16	2	4	16.00
60	7	15	3	4	16.00
59	8	17	1	7	49.00
=8				$\Sigma d = 0$	$\Sigma d^2 = 159.50$

$$\text{Co-efficient of correlation or } n = 1 - \frac{6(\Sigma d^2)}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 159.50}{8(64-1)}$$

$$= 1 - \frac{957.00}{504}$$

$$= 1 - 1.898$$

$$= -.898$$

Note. If two or more items have an equal value then their ranks average of the ranks they would have received had they differed s Thus, as above in x series 2 items are tied for 3rd rank then the rank wo $\frac{3+4}{7}$ or 3.5 and the next item would have rank of 5. Similarly, in y serie are 3 ties for 4th rank. The rank of all these items would be=5, and th item would have a rank of 7.

PROBLEM 4.26 *Ten competitors in a beauty contest are ranked b judges in the following order :*

First Judge— 1, 6, 5, 10, 3, 2, 4, 9, 7, 8,
Second Judge—3, 5, 8, 4. 7, 10, 2, 1, 6, 9,
Third Judge— 6, 4, 9, 8. 1, 2, 3, 10, 5, 7.

Use the rank correlation co-efficient to discuss which pair of judges h nearest approach to common tastes in beauty.

SOLUTION. We shall calculate the co-efficient of correlation betwe ranking of

(*a*) First and Second Judge.
(*b*) First and Third Judge.
(*c*) Second and Third Judge.

(*a*) First Judge Ranks	1	6	5	10	3	2	4	9	7	8	To
Second Judge Ranks	3	5	8	4	7	10	2	1	6	9	
Difference in Ranks (*d*)	−2	+1	—3	+6	—4	—8	+2	+8	+1	—1	$d=$
(d^2)	4	1	9	36	16	64	4	64	1	1	$\Sigma d^2=$

$$\text{Co-efficient of correlation} = 1 - \frac{6(\Sigma d^2)}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 200}{10(10^2-1)}$$

$$= 1 - \frac{1200}{990}$$

$$= 1 - 1.21$$

$$= -.21$$

(*b*) First Judge Ranks	1	6	5	10	3	2	4	9	7	8	T
Third Judge Ranks	6	4	9	8	1	2	3	10	5	7	
Difference in Ranks (*d*)	—5	+2	—4	+2	+2	0	+1	−1	+2	+1	Σ
(d^2)	25	4	16	4	4	0	1	1	4	1	Σ

Co-efficient of correlation or $r=1-\frac{6(\Sigma d^2)}{n(n^2-1)}$

$$=1-\frac{6\times 60}{10(10^2-1)}$$

$$=1-\frac{360}{990}$$

$$=1-.363$$

$$=+.637$$

(c)

cond Judge Ranks	3	5	8	4	7	10	2	1	6	9	Total
hird Judge Ranks	6	4	9	8	1	2	3	10	5	7	
ifference in Ranks	-3	$+1$	-1	-4	$+6$	$+8$	-1	-9	$+1$	$+2$	$\Sigma d=0$
(d)											
(d^2)	9	1	1	16	36	64	1	81	1	4	$\Sigma d^2=214$

Co-efficient of correlation or $r=1\ \frac{6(\Sigma d^2)}{n(n^2-1)}$

$$=1-\frac{6\times 214}{10(10^2-1)}$$

$$=1-\frac{1284}{990}$$

$$=1-1.29$$

$$=-.29$$

Thus, we find that the first and the third Judges have highest similarity of ught and have the nearest approach to common tastes in beauty.

PROBLEM 4.27 *Calculate the co-efficient of concurrent deviations from the owing data :—*

n or number of pairs of observations $=96$

c or number of pairs of concurrent deviations $=32$

SOLUTION. Substituting the given values in the formula :—

$$r=\pm\sqrt{\pm\left(\frac{2c-n}{n}\right)}$$

where r represents the co-efficient of correlation,

n, the number of pairs of observations,

and c, the number of concurrent deviations.

We get,

$$r=\pm\sqrt{\pm\left(\frac{64-96}{96}\right)}$$

$$=\pm\sqrt{-(-.333)}$$

$$=-\sqrt{.333}$$

$$=-.577$$

Thus the co-efficient of concurrent deviations of the given data is —. 574

PROBLEM 4.28 *Calculate the co-efficient of concurrent deviations fro data given below :*

Year	*Supply*	*Price*
1963	*160*	*292*
1964	*164*	*280*
1965	*172*	*260*
1966	*182*	*234*
1967	*166*	*266*
1968	*170*	*254*
1969	*178*	*230*
1970	*192*	*190*
1971	*186*	*200*

SOLUTION. Computation of co-efficient of concurrent deviation Supply and Price.

Year	Supply (x)		Price (y)		Produ
	Index of Supply	Deviation from preceding year (x)	Index of Price	Deviation from preceding year (y)	x
1963	160		292		
1964	164	+	280	—	–
1965	172	+	260	—	–
1966	182	+	234	—	–
1967	166	—	266	+	–
1968	170	+	254	—	–
1969	178	+	230	—	–
1970	192	+	190	—	–
1971	186	—	200	+	–

$n=8$, since only the years 1964 to 1971 can be used in computing co-efficient.

$c=0$, since there are no pairs having like signs.

Substituting the above in the formula

$$r=\pm\sqrt{\pm\left(\frac{2c-n}{n}\right)}$$

where r represents the co-efficient of concurrent deviations

we get,

$$r=\pm\sqrt{\pm\left(\frac{0-8}{8}\right)}$$
$$=\pm\sqrt{-(-1)}$$
$$=-\sqrt{1}$$
$$=-1.$$

Thus, as the co-efficient of correlation between supply and price by of concurrent deviations is—1 which indicates that there is perfect inverse relation between the two.

PROBLEM 4.29 *Compute the co-efficient of correlation of the following table, (by the method of concurrent deviations) relating to the marks obtained by 12 students in History and Geography respectively.*

Student	*Marks in History*	*Marks in Geography*
A	*65*	*30*
B	*40*	*55*
C	*35*	*68*
D	*75*	*28*
E	*63*	*76*
F	*80*	*25*
G	*35*	*80*
H	*20*	*85*
I	*85*	*20*
J	*65*	*35*
K	*55*	*45*
L	*33*	*65*

SOLUTION. Computation of the co-efficient of Concurrent Deviation between the marks in History and Geography by the method of concurrent deviations.

Student	Marks in History		Marks in Geography		Product
	Marks	Deviation from preceding student (x)	Marks	Deviation from preceding student (y)	(xy)
A	65		30		
B	40	—	55	+	—
C	35	—	68	+	—
D	75	+	28	—	—
E	63	—	76	+	—
F	80	+	25	—	—
G	35	—	80	+	—
H	20	—	85	+	—
I	85	+	20	—	—
J	65	—	35	+	—
K	55	—	45	+	—
L	33	—	65	+	—

Number of pairs of observations or $n=11$

Number of concurrent deviations or $c=0$

Substituting the above values in the formula

$$r=\pm\sqrt{\pm\left(\frac{2c-n}{n}\right)}$$

where r represents the co-efficient of correlation

we get,

$$r=\pm\sqrt{\pm\left(\frac{0-11}{11}\right)}$$

$$=\pm\sqrt{-(-1)}$$

$$=-\sqrt{1}$$

$$=-1.$$

Thus, the co-efficient of correlation between the marks of 12 students in History and Geography is—1, which indicates that there is perfect inverse correlation.

PROBLEM 4.30 *Compute the co-efficient of correlation, by concurrent deviation method of the following table relating to the average number of employed daily in thousands and the average number of bales consumed daily in lakhs.*

Year	*Quarters*	*Average number of employed daily (000's)*	*Average number of bales consumed daily in lakhs*
1969	*I*	*274*	*22*
	II	*275*	*25*
	III	*252*	*20*
	IV	*238*	*19*
1970	*I*	*274*	*27*
	II	*285*	*28*
	III	*293*	*30*
	IV	*290*	*29*
1971	*I*	*275*	*25*
	II	*305*	*30*
	III	*308*	*35*

SOLUTION. Computation of the co-efficient of correlation between the average number of employed daily and the average number of bales consumed daily by means of Concurrent Deviations.

Year and the Quarters		Average number employed daily		Average number of bales consumed daily		Product
		Number in (000's)	Deviation from preceding quarter (*x*)	Number of bales in lakhs	Deviation from preceding quarter (*y*)	(*xy*)
1969	I	274		22		
	II	275	+	25	+	+
	III	252	—	20	—	+
	IV	238	—	19	—	+
1970	I	274	+	27	+	+
	II	285	+	28	+	+
	III	293	+	30	+	+
	IV	290	—	29	—	+
1971	I	275	—	25	—	+
	II	305	+	30	+	+
	III	308	+	35	+	+

Number of pairs of observations or $n=10$

Number of concurrent deviations or $c=10$

Substituting the above values in the formula

$$r=\pm\sqrt{\pm\left(\frac{2c-n}{n}\right)}$$

where *r* represents the co-efficient of correlation.

we get,

$$r=\pm\sqrt{+\left(\frac{20-10}{10}\right)}$$

$$=\pm\sqrt{(1)}$$

$$=+\sqrt{1}$$

$$=+\ 1.$$

Thus, the co-efficient of correlation between the average number of employed daily and the average number of bales consumed daily is +1, which indicates that there is perfect positive correlation between the two.

PROBLEM 4.31 *Compute the co-efficient of concurrent deviations from the following table, showing the output of steel in tons and the number of unemployed persons in steel industry in thousands for 12 months.*

Months	*Output of steel in(ooo's) Tons*	*Unemployed in Steel Industry (ooo's)*
January	*8.5*	*60*
February	*9.2*	*65*
March	*9.3*	*61*
April	*8.5*	*74*
May	*7.2*	*92*
June	*5.9*	*157*
July	*5.1*	*130*
August	*6.6*	*106*
September	*7.9*	*58*
October	*7.6*	*80*
November	*8.2*	*50*
December	*9.3*	*45*

SOLUTION. Computation of the co-efficient of correlation between the output of steel and the number of unemployed persons in steel industry by means of concurrent deviations

Months	Output of steel		Unemployed in Steel Industry		Product
	Output in (ooo's) tons	Deviation from preceding month (x)	Number in(ooo's)	Deviation from preceding month (y)	(xy)
January	8.5		60		
February	9.2	+	65	+	+
March	9.3	+	61	−	−
April	8.5	−	74	+	−
May	7.2	−	92	+	−
June	5.9	−	157	+	−
July	5.1	−	130	−	+
August	6.6	+	106	−	−
September	7.9	+	58	−	−
October	7.6	−	80	+	−
November	8.2	+	50	−	−
December	9.3	+	45	−	−

Number of pairs of observations or $n=11$
Number of concurrent deviations or $c=2$
Substituting the above values in the formula

$$r=\pm\sqrt{\pm\left(\frac{2c-n}{n}\right)};$$

Where r represents the co-efficient of correlation

we get,

$$r=\pm\sqrt{\pm\left(\frac{4-11}{11}\right)}$$
$$=\pm\sqrt{-(-.6364)}$$
$$=-\sqrt{.6364}$$
$$=-.798$$

Thus, the co-efficient of correlation between the output of steel and th number of unemployed persons in steel industry is—.798, which indicates th there is a high degree of inverse correlation between the two.

PROBLEM 4.32 *Obtain the co-efficient of correlation between price of ri and rainfall from the data given below by means of concurrent deviations:—*

Year	*Price of rice in Rs. per quintal*	*Annual Rainfall in centimetres*
1959	*175.00*	*315*
1960	*160.00*	*340*
1961	*158.00*	*350*
1962	*200.00*	*350*
1963	*198.00*	*330*
1964	*195.00*	*335*
1965	*196.00*	*353*
1966	*190.00*	*333*
1967	*191.00*	*390*
1968	*195.00*	*340*
1969	*196.00*	*380*
1970	*204.00*	*340*

SOLUTION. Calculation of the co-efficient of correlation between price o rice and rainfall by means of concurrent deviations.

Year	Price of Rice		Annual Rainfall		Produc
	Price in rupees per quintal	Deviation from preceding year (x)	Rainfall inches	Deviation from preceding year (y)	(xy)
1959	175		315		
1960	160	−	340	+	−
1961	158	−	350	+	−
1962	200	+	350	0	0
1963	198	−	330	−	+
1964	195	−	335	+	−
1965	196	+	353	+	+
1966	190	−	333	−	+
1967	191	+	390	+	+
1968	195	+	340	−	−
1969	196	+	380	+	+
1970	204	+	340	−	−

Number of pairs of observations or $n=11$
Number of concurrent deviations or $c=5$

Substituting the above values in the formula

$$r=+\sqrt{+\left(\frac{2c-n}{n}\right)}$$

where r represents the co-efficient of correlation

We get,

$$r=+\sqrt{+\left(\frac{10-11}{11}\right)}$$

$$=+\sqrt{-\left(\frac{-1}{11}\right)}$$

$$=+\sqrt{.09}$$

$$=-.3$$

Thus, the co-efficient of correlation between the price of rice and rainfall is—.3

PROBLEM 4.33 *Compute the co-efficient of correlation of the short-time oscillations from the following ignoring decimals.*

Year	*Supply*	*Price*
1961	*80*	*146*
1962	*82*	*140*
1963	*86*	*130*
1964	*91*	*117*
1965	*83*	*133*
1966	*85*	*127*
1967	*89*	*115*
1968	*96*	*95*
1969	*93*	*100*

SOLUTION. Computation of co-efficient of correlation of short-time oscillations between supply and price.

Year	Supply				Price				
	Index of supply	Three year moving average	Deviation from moving average (x)	Square of deviation (x^2)	Index of price	Three year moving average	Deviation from moving average (y)	Square of Deviation (y^2)	Product of deviation (xy)
1961	80				146				
1962	82	83	—1	1	140	139	+1	1	—1
1963	86	86	0	0	130	129	+1	1	0
1964	91	87	+4	16	117	127	—10	100	—40
1965	83	86	—3	9	133	126	+7	49	—21
1966	85	86	—1	1	127	125	+2	4	—2
1967	89	90	—1	1	115	112	+3	9	—3
1968	96	93	+3	9	95	103	—8	64	—24
1969	93				100				
		$n=7$		$\Sigma x^2=$ 37		$n=7$		$\Sigma y^2=$ 228	$\Sigma xy=$ —91

Substituting the above values in Pearson's modified formula

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

where r represents the co-efficient of correlation

We get,

$$r=\frac{-91}{\sqrt{37 \times 228}}$$

$$=\frac{-91}{\sqrt{8436}}$$

$$=-.99$$

Thus the co-efficient of correlation of short-time oscillations between supply and price is —.99.

PROBLEM 4.34 *The following table gives the wholesale price index numbers for Calcutta and Kanpur for the period 1957-71 :*

Year	*Calcutta Index Nos. (base : July, 1950)*	*Kanpur Index Nos. (base : July, 1950)*
1957	*148*	*137*
1958	*145*	*137*
1959	*141*	*133*
1960	*116*	*108*
1961	*96*	*95*
1962	*91*	*99*
1963	*87*	*97*
1964	*89*	*96*
1965	*91*	*99*
1966	*91*	*102*
1967	*102*	*108*
1968	*95*	*104*
1969	*108*	*108*
1970	*120*	*116*
1971	*139*	*120*

Calculate the co-efficient of correlation of the short-time oscillations between the above time indices taking of five-yearly moving average and ignoring decimals.

SOLUTION. Calculation of the co-efficient of correlation of short-time oscillations between the Calcutta and the Kanpur Index Numbers.

Year	Calcutta Index Numbers base : July, 1950				Kanpur Index Numbers base : July, 1950				
	Index numbers	5 yearly moving average	Deviation from moving av. (x)	Square of deviation (x^2)	Index numbers	5 yearly moving average	Deviation from moving av. (y)	Square of deviation (y^2)	Product of deviations (xy)
1957	148				137				
1958	145				137				
1959	141	129	+12	144	133	122	+11	121	132
1960	116	118	−2	4	108	114	−6	36	12
1961	96	106	−10	100	95	106	−11	121	110
1962	91	96	−5	25	99	99	0	0	0
1963	87	91	−4	16	97	97	0	0	0
1964	89	90	−1	1	96	99	→ 3	9	3
1965	91	92	−1	1	99	100	−1	1	1
1966	91	94	−3	9	102	102	0	0	0
1967	102	97	+5	25	108	104	+4	16	20
1968	95	103	−8	64	104	108	−4	16	32
1969	108	113	−5	25	108	111	−3	9	15
1970	120				116				
1971	139				120				
			$\Sigma x = -22$	$\Sigma x^2 = 144$			$\Sigma y = -13$	$\Sigma y^2 = 329$	$\Sigma xy = 325$

Number of pairs of observations or $n=11$, since only the years 1959-1969 can be used in computing the co-efficient.

Substituting the above values in the Pearson's modified formula

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2+\Sigma y^2}}$$

where r represents the co-efficient of correlation

we get,

$$1=\frac{325}{\sqrt{414+329}}$$

$$=\text{Anti log}\ [\log 325-\tfrac{1}{2}\ (\log 414+\log 329)]$$
$$=\text{Anti log}\ [2.5119-\tfrac{1}{2}\ (2.6170+2.5172)]$$
$$=\text{Anti log}\ [2.5119-\tfrac{1}{2}\ (5.1342)]$$
$$=\text{Anti log}\ [2.5119-2.5671]$$
$$=\text{Anti log}\ \bar{1}.9448$$
$$=+.8806$$

PROBLEM 4.35 *Compute the co-efficient of correlation of the short-time oscillations between the cost of living and the weekly wage rates from the following data :*

Date	*Cost of living Index*	*Index of weekly wages rates*
1960	*151*	*155*
1961	*110*	*120*
1962	*102*	*99*
1963	*101*	*98*
1964	*103*	*101*
1965	*100*	*101*
1966	*100*	*102*
1967	*96*	*100*
1968	*95*	*99*
1969	*95*	*99*
1970	*87*	*98*
1971	*84*	*96*
1972	*81*	*94*

(*Assume a five-year cycle, and ignore decimals*)

SOLUTION. Computation of the co-efficient of short-time oscillations between the cost of living and the weekly wage rates.

	Cost of Living Index				Index of weekly wage rates				
Date	Index	5 yearly moving average	Deviation from moving average (x)	Square of Deviation (x^2)	Index	5 yearly moving average	Deviation from moving average (y)	Square of deviation (y^2)	Product of deviation (xy)
1960	151				155				
1961	110				120				
1962	102	113	—11	121	99	115	—16	256	176
1963	101	103	— 2	4	98	104	— 6	36	12
1964	103	101	+ 2	4	101	100	+ 1	1	2
1965	100	100	0	0	101	100	+ 1	1	0
1966	100	99	+ 1	1	102	101	+ 1	1	1
1967	96	97	— 1	1	100	100	0	0	0
1968	95	95	0	0	99	100	— 1	1	0
1969	95	91	+ 4	.16	99	98	+ 1	1	4
1970	87	88	— 1	1	98	97	+ 1	1	—1
1971	84				96				
1972	81				94				
				$\Sigma x^2=$ 148				$\Sigma y^2=$ 298	$\Sigma xy=$ 194

Number of pairs of observations $n=9$

Substituting the above values in the Pearson's modified formula

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

where r represents the co-efficient of correlation.

we get,

$$r=\frac{194}{\sqrt{148\times 298}}$$

$$=\frac{194}{\sqrt{44104}}$$

$$=\frac{194}{210}$$

$$=+.9238.$$

Thus the co-efficient of correlation of short-time oscillations between the cost of living and the weekly wage rates is $+.9238$,

PROBLEM 4.36 *Compute the co-efficient of correlation of short-time oscillations between demand and price from the following table assuming 5-yearly cycle and ignoring decimals.*

Year	*Index of demand*	*Index of price*
1957	*101*	*117*
1958	*108*	*97*
1959	*105*	*102*
1960	*145*	*118*
1961	*153*	*205*
1962	*186*	*196*
1963	*202*	*177*
1964	*207*	*168*
1965	*204*	*177*
1966	*198*	*170*
1967	*200*	*165*
1968	*208*	*170*
1969	*232*	*175*
1970	*228*	*180*
1971	*222*	*190*

(*Figures in the above table are arbitrary.*)

SOLUTION. Computation of the co-efficient of correlation of short-time oscillations between demand and price.

	Demand				Price				
Year	Index of Demand	5 yearly moving average	Deviation from moving av. (x)	Square of deviation (x^2)	Index of Price	5 yearly moving average	Deviation from moving av. (y)	Square of deviation $(y)^2$	Product of deviation (xy)
1957	101				117				
1958	108				97				
1959	105	122	—17	289	102	128	—26	676	442
1960	145	139	+ 6	36	118	144	—26	676	—156
1961	153	158	— 5	25	205	160	+45	2025	—225
1962	186	177	+ 9	81	196	173	+23	529	207
1963	202	190	+12	144	177	185	— 8	64	— 96
1964	207	199	+ 8	64	168	178	—10	100	— 80
1965	204	202	+ 2	4	177	171	+ 6	36	12
1966	198	202	— 4	16	170	170	0	0	8
1967	200	208	— 8	64	165	171	— 6	36	40
1968	208	213	— 5	25	170	172	— 2	4	10
1969	232	218	+14	196	175	176	— 1	1	—14
1970	228				180				
1971	222				190				
				$\Sigma x^2=$ 944				$\Sigma y^2=$ 4147	$\Sigma xy=$ 148

Number of pairs of observations or $n=11$

Substituting the above values in the formula.

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\times\Sigma y^2}}$$

where r represents the co-efficient of correlation

We get,

$$r=\frac{148}{\sqrt{944\times4147}}$$

$=$Anti log $[\log 148-\frac{1}{2}(\log 944+\log 4147)]$

$=$Anti log $[2.1703-\frac{1}{2}(2.9750+3.6177)]$

$=$Anti log $[2.1703-\frac{1}{2}(6.5927)]$

$=$Anti log $[2.1703-3.2964]$

$=$Anti log $\bar{2}.8731$

$=.07480$

PROBLEM 4.37 *Compute the co-efficient of correlation of short-tim oscillations between index number of Raw Cotton and Cloth from the followi table.*

Date *1974*		*Index Nnmber of Raw Cotton*	*Index Number of Cloth*
July	*7*	*96*	*104*
	14	*101*	*106*
	21	*99*	*110*
	28	*102*	*107*
August	*7*	*104*	*108*
	14	*105*	*108*
	21	*103*	*109*
	28	*102*	*109*
Sept.	*7*	*107*	*108*
	14	*106*	*107*
	21	*108*	*105*
	28	*110*	*110*
Oct.	*7*	*115*	*112*
	14	*109*	*107*
	21	*112*	*108*
	28	*107*	*108*

(Figures in the above table are arbitrary)

SOLUTION. Computation of the co-efficient of correlation of short-time oscillations between index numbers of Raw Cotton and Cloth.

Date	Index number of Raw Cotton				Index number of Cloth				
	Index number	4 weekly moving average	Deviation from moving av (*x*)	Square of deviation (*x*²)	Index number	4 weekly moving average	Deviation from moving av. (*y*)	Square of deviation (*y*²)	Product of deviation (*xy*)
1974									
July 7	96				104				
14	101				106				
21	99	101.5	—1.5	2.25	110	107.3	+2.7	7.29	— 4.05
28	102	102	0	0	107	108	—1.0	1 00	— 0
Aug.									
7	104	103	+1	1.00	108	108.1	— .1	.01	— .1
14	105	103.5	+1.5	2.25	108	108.3	— .3	.09	— .45
21	103	103.9	— .9	.81	109	108.5	+ .5	.25	— .45
28	102	104.4	—2.4	5.76	109	108.4	+ .6	.36	— 1.44
Sept.									
7	107	115.1	+1.9	3.61	108	107 8	+ .2	.04	+ .38
14	106	116.8	— .8	.64	107	107.4	— .4	.16	+ .32
21	108	108.8	— .8	.64	105	108	— .3	9.00	+ 2.40
28	110	110.1	— .1	.01	110	108.5	+1.5	2.25	— .15
Oct.									
7	115	111	+4	16.00	112	108.9	+3.1	9.61	+12.40
14	109	111.1	—2.1	4.41	107	109	—2.0	4.00	+ 4.20
21	112				108				
28	107				108				
				$\Sigma x^2=$ 37.38				$\Sigma y^2=$ 34.06	$\Sigma xy=$ 13.06

Number of pairs of observations or $n=12$.

Substituting the above values in the formula

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\times\Sigma y^2}}$$

where r represents the co-efficient of correlation

we get,

$$r=\frac{13.05}{\sqrt{37.38\times 34.06}}$$

$$r=\frac{13.05}{\sqrt{1273.1628}}$$

$$=\frac{13.05}{35.68}$$

$$=.366.$$

Thus, the co-efficient of correlation of short-time oscillations between index number of Raw Cotton and Cloth is .366.

REVISIONAL PROBLEMS-4

Karl Pearson's Co-efficient of correlation :

4.38. A sample of five items is taken from the production of a firm. The length and weight of 5 items are given below. Compute the value of Co-efficient of correlation

Length (centimetres)	3	4	6	7	10
Weigth (grams)	9	11	14	15	16

(*Ans*.+.896)

4.39. The following are the results of H.S. School examination :

Age of Candidates	Percentage of failures	Age of Candidates	Percentage of failures
13	39	18	39
14	40	19	48
15	43	20	47
16	34	21	54
17	36		

Calculate the value of r and its probable error

(*Ans* $r=+.686$; P.E=.119)

Ranking Correlation :

4.40. Following are ranks obtained by 10 students in two subjects:

Statistics :	1	2	3	4	5	6	7	8	9	10
Mathematics:	1	4	1	5	3	9	7	10	6	8

To what extent the knowledge of students in two subjects is related ?

(*Ans*: $r=+758$)

4.41. Calculate the co-efficient of rank Correlation from the following data:

x	48	33	40	9	16	16	65	24	16	57
y	13	13	24	6	15	4	20	9	9	19

(*Ans*: $r=+.533$)

Concurrent Correlation

4.42. Calculate Co-efficient of correlation by concurrent deviation **method** from the following data:

Year	Supply	Price	Year	Supply	Price
1964	150	200	1968	160	190
1965	154	180	1969	165	180
1966	160	170	1970	180	172
1967	172	160			

(*Ans*: r =−

4.43. By the methods of Concurrent Deviations and Rank Correlati calculate the Co-efficient of Correlation from the data given below:

Serial Number	1	2	3	4	5	6	7	8	9	10	11
Marks in Statistics	65	40	35	75	63	80	35	20	80	60	50
Marks in Economics	60	55	50	56	30	70	40	35	80	75	80

(*Ans*: r =+.6, r =+1.0

4.44. 100 students were examined in accountancy and statistics. They we classified according to the marks obtained by them as is shown in t following table. To what extent is the knowledge of the students in the t subjects related ?

Marks Statistics	Accountancy 20—	25—	30—	35—40
15—	20	10	2	—
25—	4	24	6	2
35—	—	5	11	1
45—	—	5	1	1
55—	—	—	—	

(*Ans*: r=+.627

4·45. Calculate the Karl Pearson's co-efficient of correlation of the sho term oscillations for the indices of supply and price of a certain commodi given below:

Year	Index of supply	Index of price	Year	Index of supply	Index of price
1960	91	117	1967	107	68
1961	98	97	1968	104	77
1962	95	102	1969	98	93
1963	92	108	1970	100	89
1964	93	105	1971	108	83
1965	96	96	1972	116	78
1966	102	77	1973	114	84
			1974	111	95

(Take 5-yearly moving average and ignore decimals in computing the average)

Hint: use $r=\frac{\Sigma xy}{\sqrt{\Sigma x^2 \times y^2}}$ (*Ans*: r=−.925

4.46. From the following data compute the co-efficient of correlation betwe x and y:

	x	y
Arithmetic Mean	25	18
Square of deviations from arithmetic mean	136	138

Summation of products of deviations of x and y series from their respective means =122

No. of pairs of values =15 (*Ans*: r=+0.89

4.47. Draw a scatter diagram to represent the following data:

x	15	18	30	27	25	23	30
y	7	10	17	16	12	13	9

Also calculate the Co-efficient of Correlation between x and y from the above data. (*Ans*: r=+0.632

Regression

5

Regression involves comparison between two or more variables. It is the basis of correlation analysis. It is a measure of average relationship between two or more series and tells about the probable change in a dependent variable as compared to a certain change in the other variable.

Unless there is a perfect correlation between two series (in which case there is only one regression equation) there are two regression equations, one of x on y and the other of y on x.

In this chapter various types of questions of regression have been solved. The method of solving the questions can easily be followed from the explanation of the ymbols used in the various problems.

PROBLEM 5.1. *Calculate Karl Pearson's co-efficient of correlation and the gression equations from the following data :—*

Age of husband	*Age of wife*
18	*17*
19	*17*
20	*18*
21	*18*
22	*18*
23	*19*
24	*19*
25	*20*
26	*21*
27	*22*

SOLUTION. Calculation of co-efficient of correlation between ages of usbands and wives.

	Age of husband (X)		Age of wife (Y)			
ge in ars m_1)	Deviation from as. av. (22) (x)	Square of deviation (x^2)	Age in years (m_2)	Deviation from as. av. (19) (y)	Square of deviation (y^2)	Product of deviations (xy)
18	—4	16	17	—2	4	8
19	—3	9	17	—2	4	6
20	—2	4	18	—1	1	2
21	—1	1	18	—1	1	1
22	0	0	18	—1	1	0
23	+1	1	19	0	0	0
24	+2	4	19	0	0	0
25	+3	9	20	+1	1	3
26	+4	16	21	+2	4	8
27	+5	25	22	+3	9	15
10	$\Sigma x=+5$	$\Sigma x^2=85$		$\Sigma y=-1$	$\Sigma y^2=25$	$\Sigma xy=43$

Substituting the above values in Pearson's formula.

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

where r represents the co-efficient of correlation.

we get,

$$r=\frac{43-10\left(\frac{5}{10}\right)\left(\frac{-1}{10}\right)}{10\sqrt{\frac{85}{10}-\left(\frac{5}{10}\right)^2}\sqrt{\frac{25}{10}-\left(\frac{-1}{10}\right)^2}}$$

$$=\frac{43\times 10+5\times 1}{\sqrt{85\times 10-(5)^2}\sqrt{25\times 10-(-1)^2}}=\frac{430+5}{\sqrt{850-25}\sqrt{250-1}}$$

$$=\frac{435}{\sqrt{825\times 249}}=.9596$$

Regression co-efficient of x on y

or $bxy=r\frac{\sigma x}{\sigma y}$

$$=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)\times\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

$$=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\left\{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2\right\}}$$

$$=\frac{\Sigma xy-\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma y^2-n\left(\frac{\Sigma y}{n}\right)^2}$$

$$= \frac{43-10\left(\frac{5}{10}\right)\left(\frac{-1}{10}\right)}{25-10\left(\frac{-1}{10}\right)^2}$$

$$= \frac{43.5}{24.9}$$

$$= 1.747$$

Similarly Regression co-efficient of y on x is represented by

$$byx = r\frac{\sigma y}{\sigma x}$$

$$= \frac{\Sigma xy - n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma x^2 - n\left(\frac{\Sigma x}{n}\right)^2} = \frac{43-10\left(\frac{5}{10}\right)\left(\frac{-1}{10}\right)}{85-10\left(\frac{5}{10}\right)^2}$$

$$= \frac{43.5}{82.5} = .527$$

Arithmetic average of X series of Ax=Assumed av.+ $\frac{\Sigma x}{n}$

$$= 22 + \frac{5}{10} = 22.5$$

Arithmetic average of Y series or Ay=As. av. + $\frac{\Sigma y}{n}$

$$= 19 + \frac{-1}{10} = 18.9$$

Substituting the above values in the following Regression Equation of x on y.

$(x-Ax) = bxy\ (y-Ay)$

we get,

$(x-22.5) = 1.747\ (y-18.9)$
or $x = 1.747y - 33.018 + 22.5$
$= 1.747y - 10.518$
Similarly, regression equation of y on x is
$(y-Ay) = byx\ (x-Ax)$
Substituting the calculated values, we get,
$(y-18.9) = .527(x-22.5)$
or $y = .527x - 11.86 + 18.9$
$= .527x + 7.04$.

Thus, the co-efficient of correlation between ages of husband and wife is .959 and the regression equations are

$x = 1.747y - 10.518$ ....(*i*)
$y = .527x + 7.04$ ....(*ii*)

PROBLEM 5.2. *Write down the two regression equations that may be associat with the following pairs of values :—*

(x) :	*152*	*114*	*138*	*154*	*144*	*153*	*141*	*117*	*136*	*154*
(y) :	*193*	*300*	*414*	*594*	*676*	*549*	*320*	*483*	*481*	*65*

SOLUTION. Calculation of the Regression Co-efficients.

Size of the item (m_1)	Deviation from as. av. (130) (x)	Square of Deviation (x^2)	Size of the item (m_2)	Deviation from as. av. (500) (y)	Square of Deviation (y^2)	Product of Deviations (xy)
152	+22	484	193	−307	94249	−6754
114	−16	256	300	−200	40000	+3200
138	+ 8	64	414	− 86	7396	− 688
154	+24	576	594	+ 94	8836	+2256
144	+14	196	676	+176	30976	+2464
153	+23	529	549	+ 49	2401	+1127
141	+11	121	320	−180	32400	−1980
117	−13	169	483	− 17	289	+ 221
136	+ 6	36	481	− 19	361	− 144
154	+24	576	659	+159	25281	+3816
	$\Sigma x=+103$	$\Sigma x^2=$ 3007		$\Sigma y=-331$	$\Sigma y^2=$ 2,42,189	$\Sigma xy=$ 3548

Regression co-efficient of x on y, or

$$bxy=r\frac{\sigma x}{\sigma y}=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma y^2-n\left(\frac{\Sigma y}{n}\right)^2}=\frac{3548-10\left(\frac{103}{10}\right)\left(\frac{-331}{10}\right)}{242189-10\left(\frac{-331}{10}\right)^2}$$

$$=\frac{3548+3409.3}{242189-10956.1}=\frac{6957.3}{231232.9}=.03$$

Similarly, the regression co-efficient of y on x

$$byx=r\frac{\sigma y}{\sigma x}=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma x^2-n\left(\frac{\Sigma x}{n}\right)}$$

$$=\frac{3548-10\left(\frac{103}{10}\right)\left(\frac{-331}{10}\right)}{3007-10\left(\frac{103}{10}\right)^2}$$

$$\frac{3548+3409.3}{3007-1060.9}=\frac{6957.3}{1946.1}=3.575$$

Arithmetic average of x series or $\dot{A}x = As.\ av: + \frac{\Sigma x}{n}$

$$=130+\frac{103}{10}=140.3$$

Arithmetic average of y series or $Ay =$ As. av. $+\frac{\Sigma y}{n}$

$$=500+\frac{-331}{10}=466.9$$

Substituting the above values in the following regression equation of x on y.

$$(x-Ax)=bxy\,(y-Ay)$$

we get,

$$(x-140.3)=.03\,(y-466.9)$$

$$\text{or } x=.03y-14.007+140.3$$

$$x=.03y+126.293$$

Similarly the regression equation of y on x is

$$(y-Ay)=byx\ \ (x-Ax)$$

Substituting the calculated values, we get

$$(y-466.9)=3.575\ (x-140.3)$$

$$\text{or } y=3.575x-501.57+466.9$$

$$y=3.575x-34.67$$

Thus the two regression equations that may be associated with the given pairs of values are :—

$$x=.03y+126.293$$

$$y=3.575\ x-34.67$$

PROBLEM 5.3. *The following marks have been obtained by a class of students in statistics (out of 100)*

Paper I	*80,*	*45,*	*55,*	*56,*	*58,*	*60,*	*65,*	*68,*	*70,*	*75,*	*85.*
Paper II	*82,*	*56,*	*50,*	*48,*	*60,*	*62,*	*64,*	*65,*	*70,*	*74,*	*90.*

Compvte the co-efficient of correlation for the above data. Find the lines of regression and examine the relationship.

SOLUTION. Calculation of the co-efficient of correlation between marks obtained by a class of students in Statistics I paper and Statistics II paper.

Paper I X			Paper II Y			
Marks out of (100) (m_1)	Deviation from as. av. (65) (x)	Square of deviation (x^2)	Marks out of (100) (m_2)	Deviation from as. av. (70) (y)	Square of Deviation (y^2)	Produc of deviation (xy)
80	+15	225	82	+12	144	180
45	−20	400	56	−14	196	280
55	−10	100	50	−20	400	200
56	− 9	81	48	−22	484	198
58	− 7	49	60	−10	100	70
60	− 5	25	62	− 8	64	40
65	0	0	64	− 6	36	0
68	+ 3	9	65	− 5	25	−15
70	+ 5	25	70	0	0	0
75	+10	100	74	+ 4	16	40
85	+20	400	90	+20	400	400
$n=11$	$\Sigma x=+2$	$\Sigma x^2=$ 1414		$\Sigma y=$ −49	$\Sigma y^2=$ 1865	$\Sigma xy=$ 1393

Substituting the above values in Karl Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

Where r represents the co-efficient of correlation.

we get,

$$r=\frac{1393-11\left(\frac{2}{11}\right)\left(\frac{-49}{11}\right)}{11\sqrt{\frac{1414}{11}-\left(\frac{2}{11}\right)^2}\sqrt{\frac{1865}{11}-\left(\frac{-49}{11}\right)^2}}$$

$$=\frac{(1393\times 11)+(2\times 49)}{\sqrt{1414\times 11-(2)^2}\sqrt{1865\times 11-(-49)^2}}$$

$$=\frac{15323+98}{\sqrt{15554-4}\sqrt{20515-2401}}=\frac{15421}{\sqrt{15550\times 18114}}=.913$$

Regression co-efficient of x on y or

$$byx=r\frac{\sigma x}{\sigma y}=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma x}{n}\right)}{\Sigma y^2-n\left(\frac{\Sigma y}{n}\right)^2}=\frac{1393-11\left(\frac{2}{11}\right)\left(\frac{-49}{11}\right)}{1865-11\left(\frac{-49}{11}\right)^2}$$

$$=\frac{1393+8.9}{1865-218.3}=\frac{1401.9}{1646.7}=.85$$

Regression co-efficient of y on x or

$$byx=r\frac{\sigma y}{\sigma x}=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)^2}{\Sigma x^2-n\left(\frac{\Sigma x}{n}\right)^2}$$

$$=\frac{1393-11\left(\frac{2}{11}\right)\left(\frac{-49}{11}\right)}{1414-11\left[\frac{2}{11}\right]^2}=\frac{1393+8.9}{1414-.356}=\frac{1401.9}{1413.64}=.99$$

Arithmetic average of the marks of the Ist Paper, or $Ax=\text{As. av.}+\frac{\Sigma x}{n}$

$$=65+\frac{2}{11}=65.2 \text{ marks.}$$

Arithmetic average of the marks of the 2nd Paper,

$$\text{or } Ay=\text{As. av.}+\frac{\Sigma y}{n}=70+\frac{-49}{11}=65.55 \text{ marks.}$$

Substituting the above values in the following regression equation of x on y

$$(x-Ax)=bxy\ (y-Ay)$$

We get,

$$(x-65.2)=.85\ (y-65.55)$$
$$\text{or } x=.85y-55.72+65.2$$
$$=.15y+9.48$$

Similarly, regression equation of y on x is

$$(y-Ay)=byx\ (x-Ax)$$

Substituting the calculated values, we get,

$$(y-65.55)=.99\ (x-65.2)$$
$$\text{or } y=.99x-64.55+65.55$$
$$=.99x+1$$

Thus, the co-efficient of correlation between the marks of the two papers is .913 and the regression equations are :

$$x=.85y+9.48$$
$$y=.99x+1$$

PROBLEM 5.4. *Explain the concepts of correlation and regression and compare their usefulness.*

Calculate the co-efficient of correlation and obtain the lines of regression for the following data :—

x	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
y	*9*	*8*	*10*	*12*	*11*	*13*	*14*	*16*	*15*

Obtain an estimate of y which should correspond on the average to x=6.2.

SOLUTION. Calculation of the co-efficient of correlation between x and y series.

	x		y			
Index number m_1	Deviation from as. av. (5) x	Square of deviation x^2	m_2	Deviation from as. av. (12) y	Square of deviation y^2	Product of deviation xy
1	-4	16	9	-3	9	$+12$
2	-3	9	8	-4	16	$+12$
3	-2	4	10	-2	4	$+4$
4	-1	1	12	0	0	0
5	0	0	11	-1	1	0
6	$+1$	1	13	$+1$	1	$+1$
7	$+2$	4	14	$+2$	4	$+4$
8	$+3$	9	16	$+4$	16	$+12$
9	$+4$	16	15	$+3$	9	$+12$
$n=9$	$\Sigma x=0$	$\Sigma x^2=60$		$\Sigma y=0$	$\Sigma y^2=60$	$\Sigma xy=57$

Substituting the values in Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\ \sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

Where r represents the co-efficient of correlation
We get,

$$r=\frac{57-9\left(\frac{0}{9}\right)\left(\frac{0}{9}\right)}{9\sqrt{\frac{60}{9}-\left(\frac{0}{9}\right)^2}\ \sqrt{\frac{60}{9}-\left(\frac{0}{9}\right)^2}}$$

$$=\frac{(57\times 9)-(0\times 0)}{\sqrt{60\times 9-(0)^2}\ \sqrt{60\times 9-(0)^2}}$$

$$=\frac{513-0}{\sqrt{540-0}\ \sqrt{540-0}}=\frac{513}{\sqrt{540\times 540}}=+.95$$

Regression Co-efficient of x on y or

$$bxy=r\ \frac{\sigma x}{\sigma y}=.95\times\sqrt{\frac{60}{9}}\div\sqrt{\frac{60}{9}}$$

$$=.95\times\frac{2.58}{2.58}=.95$$

Regression Co-efficient of y on x or

$$byx=r\ \frac{\sigma y}{\sigma x}=.95\frac{2.58}{2.58}=.95$$

Regression equation of x on y

$(x-Ax) \quad = bxy\ (y-Ay)$

or $x-5 \quad = .95\ (y-12)$

or $x \quad = .95\ y-11.4+5$

or $x \quad = .95\ y-6.4$

Regression equation of y on x

$(y-Ay) \quad = byx\ (x-Ax)$

or $y-12 = .95\ (x-5)$

or $y \quad = .95x-4.75+12$

or $y \quad = .95x+7.25$

The value of y when $x=6.2$

or $y \quad =(.95\times 6.2)+7.25$

$=5.89+7.25$

$=13.14$

PROBLEM 5.5. *Vital Statistics of the U.P.* (*in thousands*)

Years	*Fevers*	*Respiratory diseases*	*Dysentery and diarrhoea*	*Others*	*Total*
1971	*1025*	*37*	*16*	*228*	*1306*
1972	*853*	*34*	*13*	*176*	*1076*
1973	*698*	*35*	*12*	*160*	*905*
1974	*970*	*47*	*18*	*260*	*1295*

Find out 'r' of the deaths from the fevers and total deaths given above. Calculate standard error of the co-efficient and the line of regression of the deaths from fevers on total deaths.

SOLUTION. Calculation of the co-efficient of correlation between deaths from fever and total deaths.

Year	Deaths from fever (x)			Total deaths (y)			Product of deviations
	Number of deaths	Deviation from as. av. (885)	Square of deviation	Number of deaths	Deviation from as. av. (1145)	Square of deviation	
	(m_1)	(x)	(x^2)	(m_2)	(y)	(y^2)	(xy)
1971	1025	+140	19600	1306	+161	25921	22540
1972	853	— 32	1024	1076	— 69	4761	2208
1973	698	—187	34969	905	—240	57600	44880
1974	970	+ 85	7235	1295	+150	22500	12750
n=4		Σx=+6	Σx^2= 62818		Σy=+2	Σy^2= 110782	Σxy= 82378

Substituting the above values in Pearson's formula

$$r=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{n\sqrt{\frac{\Sigma x^2}{n}-\left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{\Sigma y^2}{n}-\left(\frac{\Sigma y}{n}\right)^2}}$$

Where r represents the co-efficient of correlation

We get,

$$=\frac{82378-4\left(\frac{6}{4}\right)\left(\frac{2}{4}\right)}{4\sqrt{\frac{62818}{4}-\left(\frac{6}{4}\right)^2}\sqrt{\frac{110782}{4}-\left(\frac{2}{4}\right)^2}}$$

$$=\frac{82378\times4-6\times2}{\sqrt{62818\times4-(6)^2}\sqrt{110782\times4-(2)^2}}$$

$$=\frac{329500}{\sqrt{251236\times443124}}=.987$$

Standard Error of the co-efficient of correlation

$$S.E.=\frac{1-r^2}{\sqrt{N}}\quad\frac{1-(.987)^2}{\sqrt{4.}}=\frac{1-.974}{2}$$

$$=\frac{.026}{2}=.013$$

Arithmetic average of the deaths from fever

or Ax As. av. $+\frac{\Sigma x}{n}=885+\frac{6}{4}$ = 886.5 deaths.

Arithmetic average of the total deaths.

$$Ay = \text{As. av.} + \frac{\Sigma y}{n} = 1145 + \frac{2}{4} = 1145.5 \text{ deaths.}$$

Regression co-efficient of x on y : where x represents the deaths from fevers and y the total deaths.

$$bxy = r\frac{\sigma x}{\sigma y} = \frac{\Sigma xy - n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma y^2 - n\left(\frac{\Sigma y}{n}\right)^2}$$

$$= \frac{82378 - 4\left(\frac{6}{4}\right)\left(\frac{2}{4}\right)}{110782 - 4\left(\frac{2}{4}\right)^2}$$

$$= \frac{82378 - 3}{110782 - 1} = \frac{82375}{110781} = .74$$

Substituting the above value in the regression equation of x on y

$(x - Ax) = byx\ (y - Ay)$

We get,

$(x - 886.5) = .74(y - 1145.5)$
or $x = .74y - 847.67 + 886.5$
$= .74y + 38.83.$

Thus, the co-efficient of correlation between the deaths from the fevers and total deaths is .987 and its standard error is .013. The line of regression of the deaths from fevers on total deaths is given by the following equation.

$x = .74y + 38.83.$

PROBLEM 5.6. *The height of fathers and sons are given in the following table :—*

Height of father in centimetres	*165*	*166*	*167*	*167*	*168*	*169*	*171*	*173*
Height of son in centimetres	*167*	*168*	*164*	*168*	*172*	*170*	*169*	*170*

Form the two lines of regression and calculate the expected average height of the son when the height of the father is 167.5 centimetres.

SOLUTION. Calculation of the regression co-efficient.

Height of father (x)			Height of son (y)			
Height in centimetres (m_1)	Deviation from as. av. (168) (x)	Square of deviation (x^2)	Height in centimetres (m_2)	Deviation from as. av. (168) (y)	Square of deviation (y^2)	Product o deviation (xy)
165	—3	9	167	—1	1	3
166	—2	4	168	0	0	0
167	—1	1	164	—4	16	4
167	—1	1	168	—0	0	0
168	0	0	172	+4	16	0
169	+1	1	1·70	+2	41	2
171	+3	9	169	+1	1	3
173	+5	25	170	+2	4	10
$n=8$	$\Sigma x=+2$	$\Sigma x^2=50$		$\Sigma y=+4$	$\Sigma y^2=42$	$\Sigma xy=22$

Regression co-efficient of x on y ; where x represents the height of father and y the height of son.

$$bxy=r\frac{\sigma x}{\sigma y}=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma y^2-n\left(\frac{\Sigma y}{n}\right)^2}$$

$$=\frac{22-8\left(\frac{2}{8}\right)\left(\frac{4}{8}\right)}{42-8\left(\frac{4}{8}\right)^2}$$

$$=\frac{22-1}{42-2}=\frac{21}{40}=.525$$

Regression co-efficient of y on x or

$$byx=r\frac{\sigma y}{\sigma x}=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma x^2-n\left(\frac{\Sigma x}{n}\right)^2}$$

$$=\frac{22-8\left(\frac{2}{8}\right)\left(\frac{4}{8}\right)}{50-8\left(\frac{2}{8}\right)^2}$$

$$=\frac{22-1}{50-.5}=\frac{21}{49.5}=.424$$

Arithmetic average of the heights of fathers or

$$Ax = \text{As. av.} + \frac{\Sigma x}{n} = 168 + \frac{2}{8} = 168.25 \text{ centimetres}$$

Arithmetic average of the heights of sons, or

$$Ay = \text{As av.} + \frac{\Sigma y}{n} = 168 + \frac{4}{8} = 168.5 \text{ centimetres}$$

Substituting the above values in the regression equation of x on y

$$(X - Ax) = bxy\ (Y - Ay)$$

We get,

$$\begin{aligned}(X - 168.25) &= .525\ (Y - 168.5)\\ \text{or } X &= .525y - 88.4625 + 168.25\\ &= .525y + 79.7875\end{aligned}$$

The regression equation of y on x is

$$(Y - Ay) = byx(X - Ax)$$

Substituting the calculated values, we get,

$$\begin{aligned}(Y - 168.5) &= .424\ (X - 168.25)\\ \text{or } Y &= .424\ X - 71.3380 + 168.5\\ &= .424\ X + 97.1620\end{aligned}$$

Calculating the expected average height of the son when the height of the father is 167.5 centimeters from the regression equation of y on x.

We have,

$$\begin{aligned}Y &= .424X + 97.1620\\ &= .424 \times 167.5 + 97.1620\\ &= 71.0200 + 97.1620\\ &= 168.1820\end{aligned}$$

Thus, the expected average height of the son is 168.1820 centimeters corresponding to the given height of the father.

PROBLEM 5.7. *Find the most likely price in Bombay corresponding to the price of Rs. 70 at Calcutta from the following data :—*

Average price at Calcutta	*65 ;*
at Bombay	*67.*
Standard deviation at Calcutta	*2.5 ;*
at Bombay	*3.5.*

Co-efficient of correlation is +.8 between the two prices of the commodity in the two towns.

SOLUTION. If x represents the price at Bombay, and y the price at Calcutta, the likely price in Bombay corresponding to the given price at Calcutta will have to be calculated from the regression equation of x on y which is,

$$(X - Ax) = r\ \frac{\sigma x}{\sigma y}\ (Y - Ay)$$

where Ax and Ay represent the average prices at Bombay and Calcutta respectively and σx and σy, the standard deviation at Bombay and Calcutta and r, the co-efficient of correlation between x and y.

Substituting the given values of Ax, Ay, σx, σy and r, we get,

$$(X - 67) = .8\ \frac{3.5}{2.5}(y - 65)$$

$$\text{or } X = 1.12\,y - 72.8 + 67$$
$$= 1.12\,y - 5.8 \quad \ldots \quad \ldots \quad (i)$$

Calculating the value of x, when y is equal to 70 from ..

We get,

$$X70 = 1.12 \times 70 - 5.8 = 78.4 - 5.8 = 72.6$$

Thus, the most likely price in Bombay corresponding to the price of Rs. 70 at Calcutta is Rs. 72.6.

PROBLEM 5.8. *Explain what is meant by a scatter diagram and line of regression. Why should there be, in general two lines of regression for each bivariate distribution ?*

You are given the following results for the heights (x) *and weight* (y) *of 1,000 policement of U.P. :—*

$Ax = 168.00$	$Ay = 70$ kg. $r = 0.60$
$\sigma x = 7.00$ cm.	$\sigma y = 10.00$ kg.

Estimate from the above data (*a*) *the height of a particular policeman whose weight is 80 kg.* (*b*) *the weight of a particular policeman who is 160 cm. tall.*

SOLUTION. If x represents the height and y the weight, the height and weight of particular policemen for the given weight and height will have to be calculated from the regression equation of x on y and of y on x respectively. They are :—

$$(x - Ax) = r\,\frac{\sigma x}{\sigma y}(y - Ay) \quad \ldots \quad \ldots \quad \ldots \quad (i)$$

$$\text{and } (y - Ay) = r\,\frac{\sigma y}{\sigma x}(x - Ax) \quad \ldots \quad \ldots \quad \ldots \quad (ii)$$

Substituting the given values of Ax, Ay, σx, σy, and r in (i) and (ii), we get,

$$(x - 168) = .60\,\frac{7}{10.00}(y - 70)$$

$$\text{or } x = .42y - 29.4 + 168$$
$$= .42y + 138.6 \quad \ldots \quad \ldots \quad \ldots \quad (iii)$$

$$\text{and } (y - 70) = .6\,\frac{10.00}{7}(x - 168)$$

$$\text{or } y = .857x - 144.00 + 70$$
$$= .857x - 74 \quad \ldots \quad \ldots \quad \ldots \quad (iv)$$

Calculating the value of x when y is equal to 80 kg. from (iii) and the value of y when x is equal to 160 cm. from (iv) :

We get,

$$X80 = .42 \times 80 + 138.6 = 33.6 + 138.6 = 172.2 \text{ cm.}$$
$$\text{and } Y\,160 = .857 \times 160 - 74 = 137.120 - 74 = 63.120 \text{ kg.}$$

Thus (i) the height of a particular policeman whose weight is 80 kg. is 172.20 cm. and,

(ii) the weight of particular policeman who is 160 cm. tall is 63.120 kg.

PROBLEM 5.9. *A study of gram prices at Hapur and Varanasi yields the following data :—*

	Hapur	*Varanasi*
	Rs.	*Rs.*
Average price per kg.	*2.463*	*2.797*
Standard deviation	*.326*	*.207*
r between prices at Hapur and Varanasi	*+0.774*	

Estimate from the above data the most likely price of gram (1) at Hapur corresponding to the price of Rs. 2.334 per kg. at Varanasi ; and (2) at Varanasi corresponding to the price of Rs. 3.052 per kg. at Hapur.

SOLUTION. If x represents the gram prices at Hapur and y at Varanasi, the most likely price of wheat at Hapur corresponding to the given price at Varanasi and the most likely price of gram at Varanasi for a given price at Hapur will have to be calculated from the regression equations of x on y and of y on x respectively. They are :

$$(x-Ax)=r\frac{\sigma x}{\sigma y}(y-Ay) \quad \ldots \quad \ldots \quad (i)$$

$$\text{and } (y-Ay)=r\frac{\sigma y}{\sigma x}(x-Ax) \quad \ldots \quad \ldots \quad (ii)$$

Substituting the given values of Ax, Ay, σx, σy and r in (i) and (ii), we get,

$$(x-2.463)=.774\frac{.326}{.207}(y-2.797)$$

$$\begin{aligned}\text{or} \quad x&=1.2189y-3.409+2.463\\ &=1.2189y-.946 \quad \ldots \quad \ldots \quad (iii)\end{aligned}$$

$$\text{and } (y-2.797)=\cdot774\,\frac{.207}{.326}(x-2.463)$$

$$\begin{aligned}\text{or } y&=.491x-1.210+2.797\\ &=.491x+1.587 \quad \ldots \quad \ldots \quad \ldots \quad (iv)\end{aligned}$$

Calculating the value of x when y is equal to 2.334 from the (iii) and the value of y when x is equal to 3.052 from the (iv) ;

We get,

$$\begin{aligned}X2.334&=1.2189\times2.334-.946\\ &=2.845-.946=1.899\end{aligned}$$

$$\begin{aligned}\text{and } Y3.052&=.491\times3.052+1.587\\ &=1.498+1.586=3.084\end{aligned}$$

Thus, the most likely price of wheat

(1) at Hapur corresponding to the price of Rs. 2.334 per kg. at Varanasi is Rs. 1.899 per kg. and

(2) at Varanasi corresponding to the price of Rs. 3.052 per kg. at Hapur is Rs. 3.084 per kg.

PROBLEM 5.10. *The following data are given for marks in English and Maths. in the S. L. C. exam. of the State in a certain year :—*

Mean marks in English =39.5
Mean marks in Maths. =47.6
Standard deviation of marks in English =10.8

Standard deviation of marks in Maths. $=16.9$
r between marks in English and Maths. $=+0{\cdot}42$

Form the two lines of regression and explain why there are two equation regression. Calculate the expected average marks in Maths of candidates received 50 marks in English.

SOLUTION. If x represents the marks in English and y that in Mathema the two lines of regression in the form of equations may be given as follows :—

$$(x-Ax)=r\ \frac{\sigma x}{\sigma y}\ (y-Ay) \ldots\ldots\ldots\ldots (i)$$

$$\text{and } (y-Ay)=r\ \frac{\sigma y}{\sigma x}\ (x-Ax) \ldots\ldots\ldots\ldots (ii)$$

Where Ax represents the mean marks in English,
Ay, the mean marks in Mathematics,
σx, the standard deviation of marks in English,
σy, the standard deviation of marks in Mathematics,

and r, the co-efficient of correlation between marks in English Mathematics,

Substituting the given values of Ax, Ay, σx, σy and r in (i) and (ii) ; we get

$$(x-39.5)=.42\ \frac{10.8}{16.9}(y-47.6)$$

$$\text{or } x.268y-12.77+39.5=.268y+26.73 \ldots\ldots \quad \ldots \quad \ldots\ldots (iii)$$

$$\text{and } (y-47.6)=.42\ \frac{16.9}{10.8}\ (x-39.5)$$

$$\text{or } y.657x-25.96+47.6=.657x+21.64 \ldots \quad \ldots \quad \ldots \quad \ldots\ldots (iv)$$

Calculating the value of y when x is equal to 50 from (iv) ; we get

$$Y50=.657\times 50+21.64$$
$$=32.85+21.64=54.49$$

Thus the two lines of regression are
$x=.268y+26.73$ and $y=.657x+21.64$

and the expected average marks in Mathematics of candidates who secured 50 mar in English are 54.49 or 54.5

PROBLEM 5.11. *In the following table are recorded data showing the scores made by salesmen on an intelligence test and their weekly sales :—*

Salesman ...	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Test Scores ..	40	70	50	60	80	50	90	40	60
Sales ('000)...	2.5	6.0	4.5	5.0	4.5	2.0	5.5	3.0	4.5

Calculate the regression line of sales on test score, and estimate the probable weekly sales volume if a salesman makes a score of 70, what will be sampling error of your estimate ?

SOLUTION. Calculation of the regression line of sales on test score.

Salesman	Test Scores (x)			Sales (y)			Product of deviation
	Scores (m_1)	Deviation from the as. av. (60) (x)	Square of deviation (x)	Sales in (000) (m_2)	Deviation from the as. av. (4.5) (y)	Square of deviation (y^2)	(xy)
1	40	—20	400	2.5	−2.0	4.00	40
2	70	+10	100	6.0	+1.5	2.25	15
3	50	—10	100	4.5	0	0	0
4	60	0	0	5.0	+ .5	.25	0
5	80	+20	400	4.5	0	0	0
6	50	—10	100	2.0	—2.5	6.25	25
7	90	+30	900	5.5	+1.0	1.00	30
8	40	—20	400	3.0	—1.5	2.25	30
9	60	0	0	4.5	0	0	0
10	60	0	0	3.0	—1.5	2.25	0
$n=10$		$\Sigma x=0$	$\Sigma x^2=$ 2400		$\Sigma y=$ —4.5	$\Sigma y^2=$ 18.25	$\Sigma xy=$ 140

Regression co-efficient of y on x ; where y represents the sales and x the test scores.

$$byx=r\ \frac{\sigma y}{\sigma x}=\frac{\Sigma xy-n\left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\Sigma x^2-n\left(\frac{\Sigma x}{n}\right)^2}$$

$$=\frac{140-10\left(\frac{0}{10}\right)\left(\frac{-4.5}{10}\right)}{2400-10\left(\frac{0}{10}\right)^2}$$

$$=\frac{140+0}{2400-0}=\frac{140}{2400}=.058=.06 \text{ apprcx.}$$

Arithmetic average of the test scores or Ax+As. av.+ $\frac{\Sigma x}{n}$

$$60+=\frac{0}{10}=60$$

Arithmetic average of the sales (000) or Ay=As. av.+ $\frac{\Sigma y}{n}$

$$4.5+=\frac{-4.5}{10}=4.05$$

The regression line of y on x is $(y-Ay)=byx(x-Ax)$

Substituting the calculated values of Ay, byx and Ax, we get

$(y-4.05)=.06\ (x-60)$
or $y=.06x-3.60+4.05=.06x+.45$ ... ... (*i*)

Calculating the value of y when x is equal to 70 from the (*i*) we get,

$y70=.06\times70+.45=4.2+.45=4.65$

Sampling Error of the above estimate

Standard Error of the regression Co-efficient of y on $x=$

$$\sqrt{\frac{\Sigma y^2-(byx\times\Sigma xy)}{n}}$$

Where Σy^2 stands for the total of the square of deviations from the actual arithmetic average and $x\ y$ for the total of the product of deviations from the actual arithmetic average. Substituting the values we get,

$$\Sigma y^2=18.25-10\left(\frac{-4.5}{10}\right)=16.225$$

$\Sigma xy=140$

or Standard Error of the regression Co-efficient of y on x

$$=\sqrt{\frac{16.225-(.06\times140)}{10}}=.88$$

$\therefore$ The Sampling Error of the above estimate (of $y_{72}=4.65$) at one standard error$=.88$

$\therefore$ The confidence limits are $4.65\pm.88$

Thus the regression line of sales on test score is $y=.06x+.45$ and the most probable weekly sales volume for a given score of 70 is 4.65 with a sampling error of .88 (at one standard error). At two Standard Error it would be 1.76 and the confidence limits 4.65 ± 1.76

PROBLEM 5.12 *Given*

$x=.85y$; $y=.89x$: $\sigma x=3$;

Find out :—

(*a*) σy (*b*) r

SOLUTION. From the given data.

Regression of x on y or $bxy=r\dfrac{\sigma x}{\sigma y}=.85$

Regression of y on x or $byx=r\ \dfrac{\sigma y}{\sigma x}=.89$

Substituting these values in the following formula

$r=\sqrt{bxy\times byx}$

we get,

$r=\sqrt{.85\times.89}=\sqrt{.7565}=\sqrt{.8697}=.87$ approx.

Similarly, $r\ \dfrac{\sigma x}{\sigma y}=.85$

or.85 $\sigma y = r \times \sigma x$

or $\sigma y = \frac{r \times \sigma x}{.85}$; substituting the values of r and σx

we get, $\sigma y = \frac{.87 \times 3}{.85} = 3.07$

Thus $r = .87$ and $\sigma y = 3.07$

PROBLEM 5.13. *In a partially destroyed laboratory record an analysis of correlation data the following results only are legible :—*

Variance of $x = 9$

Regression Equations :—

$$8x - 10y + 66 = 0.$$
$$40x - 18y \quad = 214.$$

What were (a) the mean value of x and y, (b) the standard deviation of y, and (c) the co-efficient of correlation between x and y ?

SOLUTION. (*a*) *Calculation of the mean values of x and y*

$$8x - 10y = -66 \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots(i)$$
$$40x - 18y = 214 \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots(ii)$$

Multiplying the (*i*) by 5, we have

$$40x - 18y = 214 \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots(ii)$$
$$40x - 50y = -330 \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots(iii)$$
$$32y = 544 \text{ or } y = 17.$$

Similarly, multiplying the second equation by 10 and (*i*) by 18, we have,

$$400x - 180y = 2140 \ldots \quad \ldots \quad \ldots(iv)$$
$$\underline{144x - 180y = -1188 \ldots \quad \ldots \quad \ldots(v)} \text{ or } x = 13$$
$$256x = 3328$$

(*b*) *Calculation of the co-efficient of correlation between x and y.*

The given equations are :—

$$40x - 18y \quad = 214 \ldots \quad \ldots \quad \ldots \quad \ldots(i)$$
$$8x - 10y + 66 = 0 \ldots \quad \ldots \quad \ldots \quad \ldots(ii)$$

from the first,

$$40x = 18y + 214 \text{ or } x = .45y + 5.35.$$

Therefore the regression of x on y or $bxy = .45$
Similarly, from the second

$$10y = 8x + 66$$
$$\text{or } y = .8x + 6.6$$

Therefore the regression of y on x or $byx = .8$

Now,

the co-efficient of correlation between x and y or $r = \sqrt{bxy \times byx}$

Substituting the values of bxy and byx ; we get

$$r = \sqrt{.45 \times .8} = \sqrt{.36} = .6$$

Thus, the co-efficient of correlation between x and y is .6

(c) *Calculation of the standard deviation of y.*

Regression of x on $y = r \ \dfrac{\sigma x}{\sigma y} = .45$

$$\text{or } \sigma y = \frac{r \times \sigma x}{.45}$$

Substituting the values of r and σx, we get $\sigma y = \dfrac{.6 \times 3}{.45} = 4$

Thus,

(*a*) the mean values of x and y are 13 and 17 respectively.

(*b*) the co-efficient of correlation between x and y is .6

and (*c*) the standard deviation of y is 4.

Note. In calculating the regression co-efficients if we presume that the first equation (*i.e.* $8x-10y+66=0$) is the regression equation of x on y and the other one ($40x-18y=214$) is the regression equation of y on x then the value of "r" would be more than unity. Therefore we have to presume the other way, *i.e.*, the first one as the regression equation of y on x and the other one of x on y.

REVISIONAL PROBLEMS –5

5.14. Obtain the two regression lines for the following data :

X	1	2	3	4	5	6	7	8	9
Y	9	9	10	12	11	13	14	16	15

Obtain the estimate of y which should correspond on the average to $x=6.2$

(*Ans.* $r=.95$
$x=.95y-6.4$
$y=.95x+7.25$
When $x=6.2$, $y=13.14$)

5.15. The following data give the hardness (x) and tensile strength (y) of 7 samples of metal in certain units. Find the linear regression equation of y on x.

X	146	152	158	164	170	176	182
Y	75	78	77	79	82	85	86

(*Ans.* $y=.31x+29.46$)

5.16. Calculate the co-efficient of correlation and obtain the lines of regres-ion for the following data :

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

btain the estimate of y which should correspon d on the average to $x=6.2$,

(*Ans.* $y=5.76+7.284x$
$x=.929y-6.219$
$y_{6\cdot2}=13.0438$)

5.17. Following table gives the ages of husbands and wives for 50 newly arried couples. Find the two regression lines. Also estimate :

(a) the age of husband when wife is 20, and

(b) the age of wife when husband is 30.

Age of wife	Age of husband			
	20—25	25—30	30—35	Total
16—20	9	14	—	23
20—24	6	11	3	20
24—28	—	—	7	7
Total	15	25	10	50

(*Ans.* $X=.76y+11.25$; $y=14.7+7.49x$
(*a*) *26.45 years* (*b*) *22.19 years.*)

5.18. Compute the least squares regression of *y* on *x* from the following data :

X	89	86	74	65	64	63	66	67	72	79
Y	92	91	84	75	73	72	71	75	78	84

(*Ans.* $y=0.808x+20.92$)

5.19. Given the following values, find the expected value of *x* when y is 12 :

Average of *x* Series=25
Average of y Series=22
Standard Deviation of *x* Series=4
Standard Deviation of y Series=5
Co-efficient of Correlation=0.8

(*Ans.* $x=.64y+19.92$
$x_{12}=18.6$)

5.20. From the data given below estimate the most likely height of a father whose son's height is 170 centimetres :

Father's mean height is 167 centimetres with Standard Deviation of 7 centimetres. Son's mean height is 165 centimetres with a Standard Deviation of 6 centimetres. Co-efficient of correlation between the height of father and son is+.8.

(*Ans.* *171.10*)

5.21. Two random variables have the least square regression lines with quations $3x+2y-26=0$ and $6x+y-31=0$. Find the mean values and co-fficient of correlation between *x* and y.

(*Ans.* $\bar{x}=4$; $\bar{y}=7, r=-.5$)

5.22. You are supplied with the following data only :

Variance of $x=36$
$12x-15y+99=0$
$64x-27y=321$

Calculate : (a) the average value of *x* and *y*,
(b) the standard deviation of *y*, and
(c) co-efficient of correlation between *x* and *y*.

(*Ans,* $\bar{x}=13$; $\bar{y}=17$; $r=0.6$, $\sigma y=8$)

5.23. The following calculations have been made for closing prices of 12 stocks (x) on the Bombay Stock Exchange on a certain day, alongwith the volume of sales in thousands of shares (y). From these calculations find the regression equations—

$\Sigma x=580$; $\Sigma y=370$; $\Sigma xy=11{,}494$
$\Sigma x^2=41{,}658$; $\Sigma y^2=17206$

(*Ans.* :
$y = 53.55 - 0.47x$
$x = 79.16 - 1.1y$)

5.24. The regression equations calculated from a given set of observat
are :

$$x = -.2y + 4.2$$
$$y = -.8x + 8.4$$

Calculate :

(a) $\overline{x}$ and $\overline{y}$ (b) r, (c) the estimated value of y when $x=$

(*Ans* : $\overline{x} = 3$; $\overline{y} = 6$; (*ii*) $r = -.4$; (*iii*) 5.

5.25. For 50 students of a class the equation of marks in statistics (x
marks in Accountancy (y) is $3y - 5x + 180 = 0$. The mean marks in Accounta
is 44 and variance of marks in statistics is 9/16th of the variance of mark
Accountancy. Find the mean marks in statistics and the co-efficient of correla
between marks in two subjects.

(*Ans* ·: $\overline{x} = 62.4$; $r =$

Index Numbers

6

Index numbers indicate the level of a phenomenon on a particular date or during a specific period, in comparison to the level of the same phenomenon on the previous date or period.

In this chapter various types of questions connected with the construction of index numbers have been solved. In the beginning, questions on unweighted or simple index numbers have been solved and afterwards questions on weighted index numbers, including the Fisher's Ideal index number, have been done.

PROBLEM 6.1 *Compute the index numbers for each year from the following average annual wholesale price of gur in Hapur in rupees per bag of 80 kgs. for the period 1958 to 1974 :*

Year	*Rupees*	*Year*	*Rupees*
1958	*78*	*1966*	*88*
1959	*54*	*1967*	*78*
1960	*67*	*1968*	*76*
1961	*56*	*1969*	*112*
1962	*72*	*1970*	*99*
1963	*102*	*1971*	*76*
1964	*98*	*1972*	*75*
1965	*94*	*1973*	*71*
		1974	*100*

SOLUTION. Hapur Index Numbers of gur prices.

(Base: 1958=100)

Year	Price in rupees	Index No. $\frac{C.P}{B.C.}\times 100$	Year	Price in rupees	Index Nos. $\frac{\text{Current year's price}}{\text{Base year's price}}\times 100$
1958	78	100.00	1967	78	100.00
1959	54	69.23	1968	76	97.43
1960	67	85.89	1969	112	143.58
1961	56	71.79	1970	99	126.92
1962	72	92.30	1971	76	97.43
1963	102	130.76	1972	75	96.15
1964	98	125.64	1973	71	91.02
1965	94	120.51	1974	100	128.20
1966	88	112.82			

PROBLEM 6.2 *The following data give the number of wage earners in manufactories in the U. S. and their total yearly wages. Make an Index comparing wages with 1963 as a base:—*

Year	*Wage Earners*	*Total Wages*
1954	*6,896,190*	*4,067,718,700*
1959	*9,000,059*	*10,461,786,000*
1961	*6,946,570*	*8,202,324,300*
1963	*8,778,156*	*11,009,297,000*
1965	*8,384,261*	*10,729,968,000*
1967	*8,349,755*	*10,848,802,000*
1969	*8,838,743*	*11,620,973,000*

SOLUTION. Index numbers of wages in the manufactories of the United States.

(Base 1963=100)

Year	Wage-Earner	Total-wages	Wages per head	Index Number
1954	6,896,190	$4,067,718,700	$589.85	47.03
1959	9,000,059	10,461,786,000	1162.41	92.68
1961	6,946,570	8,202,324,300	1180.77	94.14
1963	8,778,156	11,009,297,000	1254.16	100.00
1965	8,384,261	10,729,968,000	1279.77	102.04
1967	8,349,755	10,848,802,000	1299.29	103.59
1969	8,838,743	11,620,973,000	1314.77	104.83

PROBLEM 6.3 *Construct appropriate index numbers to discuss the fluctuations in the export of raw cotton, and raw jute from India for the period 1970—71 to 1975—76, using the average of the 1966—70 as base.*

Year	*Raw Cotton Quantity (1000 tons)*	*Cotton Value (lakhs of Rs.)*	*Raw Jute Quantity (1000 tons)*	*Jute Value (lakhs of Rs.)*
1966—70 (average)	*609*	*5941*	*826*	*2924*
1970—71	*701*	*4633*	*620*	*1288*
1971—72	*423*	*2345*	*587*	*1119*
1972—73	*365*	*2037*	*563*	*973*
1973—74	*504*	*2753*	*748*	*1093*
1974—75	*623*	*3495*	*752*	*1087*
1975—76	*607*	*3377*	*771*	*1371*

SOLUTION. Index Numbers of the export of Cotton and Raw Jute from India (Base : 1966—70=100)

Year	Raw Cotton				Raw Jute			
	Quantity (1000 tons)	Index Numbers	Value (lakhs of Rs.)	Index Numbers)	Quantity (1000 tons)	Index Numbers	Value (lakhs of Rs.)	Index Numbers
1966—70 (average)	609	100.00	5941	100.00	826	100.00	2924	100.00
1970—71	701	115.10	4633	77.98	620	75.06	1288	44.04
1971—72	423	69.46	2345	39.47	587	71.06	1119	38.26
1972—73	365	59.93	2037	34.28	563	68.15	973	33.27
1973—74	504	82.75	2753	46.34	748	90.55	1093	37.38
1974—75	623	102.29	3495	58.82	752	91 04	1087	37.17
1973—76	607	99.67	3377	56.84	771	93 34	1371	46.88

PROBLEM 6.4 *Use the following data of industrial production in India o compare the annual fluctuations in Indian industrial activity by the chain base ethod :*

Index Number of Industrial Production in India.

Year	*Index No.*	*Year*	*Index No.*
1959—60	*120*	*1966—67*	*149*
—61	*122*	*—68*	*156*
—62	*116*	*—69*	*137*
—63	*120*	*—70*	*162*
—64	*120*	*—71*	*149*
—65	*137*	*—72*	*160*
—66	*136*	*—73*	*160*

SOLUTION. Computation of Chain Base Index Numbers of Industrial Production in India.

Year	Index No.	Chain Base Index Number	Year	Index No.	Chain Base Index Number
1959—60	120	100.00	1966—67	149	109.55
1960—61	122	101.66	1967—68	156	104.69
1961—62	116	95.08	1968—69	137	87.82
1962—63	120	104.44	1969—70	162	118.24
1963—64	120	100.00	1970—71	149	91.97
1964—65	137	114.16	1971—72	160	107.38
1965—66	136	99.27	1972—73	160	100.00

PROBLEM 6.5 *The following table gives the average wholesale prices of the commodities A, B and C during the years 1964 to 1971.*

Average wholesale prices in Rupees

Commodities	*1964*	*'65*	*'66*	*'67*	*'68*	*'69*	*'70*	*'71*
1. A	*50.6*	*61.6*	*66.8*	*71.0*	*70.6*	*72.0*	*72.0*	*75.6*
2. B	*6.8*	*6.4*	*5.6*	*6.2*	*6.4*	*7.8*	*6.0*	*6.8*
3. C	*29.6*	*25.8*	*26.4*	*28.6*	*28.6*	*30.2*	*28.0*	*34.6*

Find out the index numbers

(i) by reference to 1964 as base year

(ii) by the chain base method.

SOLUTION.

(*i*) Index Numbers of Prices.

(Fixed Base : 1964=100)

(Percentages or relatives with base 1964=100)

Commodities	1964	'65	'66	'67	'68	'69	'70	'71
1. A	100	121.7	132.0	140.3	139.5	142.3	142.3	149.4
2. B	100	94.1	82.4	91.2	94.1	114.7	88.2	100.0
3. C	100	87.2	89.2	96.6	96.6	102.0	94.6	116.8
Total of Relatives	300	303.0	306.3	328.1	330.2	359.0	325.1	366.2
Index number as average of relatives	100	101.00	101.20	109.36	110.06	119.66	108.36	122.06

(*ii*) Index numbers of Prices.

(Chain Base Method)

(1964=100)

Percentage based on preceding year

Commodities	1964	'65	'66	'67	'68	'69	'70	'71
1. A. ..	100	121.7	108.4	106.3	99.4	102.0	100.0	10
2. B. ..	100	94.1	87.5	110.7	103.5	121.9	76.9	11
3. C. ..	100	87.2	102.3	108.3	100.0	105.6	92.9	12
Total of Link Relatives	300	303.0	298.2	325.3	302.6	329.5	269.8	33
Average Link Relatives	100	101.0	99.4	108.4	100.9	109.8	89.9	11
Chain-indices (1964=100)	100	101.0	100.4	101.6	102.5	112.6	102.1	11

PROBLEM 6.6 *You are given the following series of index numbers wholesale prices of four commodities and a straight index based on the averag Calculate a new index for the five years based on the chain method.*

Index-Numbers of wholesale prices of commodities

Year	*A*	*B*	*C*	*D*	*Total*	*Average*
1966	*308*	*476*	*220*	*324*	*1328*	*332*
1967	*216*	*512*	*328*	*248*	*1304*	*326*
1968	*348*	*444*	*400*	*416*	*1608*	*402*
1969	*300*	*616*	*384*	*372*	*1672*	*418*
1970	*172*	*660*	*352*	*240*	*1424*	*356*
1971	*176*	*636*	*356*	*240*	*1408*	*352*

SOLUTION.

Computation of Index Numbers of wholesale prices based on Chain Bas Method.

Commodities	1966	1967		1968		1969		1970		1971	
	Given Index numbers	Given Index Nos.	Percentage based on 1966	Given Index Nos.	Percentage based on 1967	Given Index Nos.	Percentage based on 1968	Given Index Nos.	Percentage based on 1969	Given Index Nos.	Percentage based on 1970
1. A	308	216	70	348	161	300	86	172	57	176	10
2. B	476	512	108	444	87	616	139	660	107	636	9
3. C	220	328	149	400	122	384	96	352	92	356	10
4. D	324	248	76	416	168	372	89	240	65	240	10
Total of Relatives	1328	1304	403	1608	538	1672	410	1424	321	1408	39
Average	332	326	100.75	402	134.5	418	102.5	356	80.25	352	99.7
Chain Indices	332		334.49		449.88		461.12		370.04		369.1

PROBLEM 6.7

(a) From the fixed base index numbers given below, prepare chain base index numbers :—

1968	*1969*	*1970*	*1971*	*1972*	*1973*
376	*392*	*408*	*380*	*392*	*400*

(b) From the chain base index numbers given below, prepare fixed base index numbers.

1968	*1969*	*1970*	*1971*	*1972*	*1973*
92	*102*	*104*	*98*	*103*	*101*

SOLUTION. (*a*) Computation of Chain Base Index Numbers from the given fixed base index numbers.

Year (1)	Fixed Base Index Numbers (2)	Fixed base index numbers changed to chain base index numbers (3)	Chain Base Index Numbers (4)
1968	376		100
1969	392	$\frac{392}{376} \times 100$	104.25
1970	408	$\frac{408}{392} \times 100$	104.08
1971	380	$\frac{380}{408} \times 100$	93.13
1972	392	$\frac{392}{380} \times 100$	103.15
1973	400	$\frac{400}{392} \times 100$	102.04

(*b*) By chaining together the chain base index numbers, we get the fixed base index numbers.

Computation of fixed base index numbers from the given chain base index numbers.

Year	Chain base index numbers	Chain base index numbers chained to 1968 as base	Fixed base index numbers
1968	92		92
1969	102	$\frac{92}{100} \times 102$	93.84
1970	104	$\frac{92}{100} \times \frac{102}{100} \times 104$	97.59
1971	98	$\frac{92}{100} \times \frac{102}{100} \times \frac{104}{100} \times 98$	95.63
1972	103	$\frac{92}{100} \times \frac{102}{101} \times \frac{104}{100} \times \frac{98}{100} \times 103$	98.50
1973	101	$\frac{92}{100} \times \frac{102}{100} \times \frac{104}{100} \times \frac{98}{100} \times \frac{103}{100} \times 101$	99.48

PROBLEM 6.8 *Which average would you use in computing the price In Number from the following data for 1974 on the basis of 1970 ? Give your reasons.*

Commodity	*Unit*	*Price in 1970*		*Price in 197*	
		Rs.	*P.*	*Rs.*	*P.*
1. Rice	*per Quintal*	*234*	*50*	*270*	*20*
2, Wheat	*per ,,*	*135*	*60*	*180*	*80*
3. Linseed	*per ,,*	*160*	*80*	*240*	*40*
4. Gur	*per ,,*	*106*	*40*	*136*	*40*
5. Cotton	*per ,,*	*1117*	*40*	*1112*	*50*
6. Tobacco	*per ,,*	*815*	*00*	*911*	*40*

SOLUTION. Index Numbers of wholesale prices of certain commodities. (1970=100)

Commodity	Unit	1970			1974		
		Prices Rs.	P.	Percentages	Prices Rs.	P.	Percentages based on 1970 a base yea
1. Rice	per qtl.	234	50	100	270	20	115
2. Wheat	,, ,,	135	60	100	180	80	133
3. Linseed	,, ,,	160	80	100	240	40	150
4. Gur	,, ,,	106	40	100	136	40	128
5. Cotton	,, ,,	1117	40	100	1112	50	100
6. Tobacco	,, ,,	815	00	100	911	40	111
Total of percentage				600			737
Index Numbers expressed as Average of percentages				100			123
Index Numbers expressed as Median of percentages				100			121.5
Index Numbers expressed as geometric mean of percentages				100			121.8

Note:—Reasons can be given in support of geometric Mean.

PROBLEM 6.9 *The following table shows the index numbers of wholesal prices of certain commodities in 1962 and 1972 (July, 1960, being taken as 100). Discuss critically how you would compare the average ratio of prices in 1972 t those in 1962, commenting on the relative advantages and disadvantages of alterna tive methods which may be used for this purpose :—*

Commodity	*Index Number of Prices*	
	1962	*1972*
Jute	*93*	*56*
Jute manufactures	*146*	*67*
Cotton	*167*	*89*
Cotton manufactures	*159*	*117*
Wool and silk	*126*	*126*

SOLUTION. Computation of the General Index Numbers of wholesale prices in 1962 and 1972 (July, 1960=100)

Commodities	Index No. of prices	
	1962	1972
1. Jute raw	93	56
2. Jute manufactures	146	67
3. Cotton raw	167	89
4. Cotton manufactures	159	117
5. Wool and siik	126	126
Total of relatives	691	455
Average of relatives	138.2	91
Median of relatives	146	89
Geometric mean of relatives	135.4	86.8

A note can be given here in support of the geometric mean.

PROBLEM 6.10 *From the following data, prepare a weighted index number for the food group for 1969 with 1959 as the base period.*

Items in the food group	*Weights*	*Price per kg. in 1959*		*Price per kg. in 1969*	
		Rs.	*ps.*	*Rs.*	*ps.*
1. Wheat	*40*	*0*	*40*	*1*	*20*
2. Rice	*20*	*0*	*80*	*2*	*00*
3. Gram	*15*	*1*	*00*	*1*	*60*
4. Arhar dal	*5*	*1*	*30*	*1*	*95*
5. Milk	*6*	*0*	*60*	*1*	*20*
6. Mustard oil	*10*	*2*	*00*	*4*	*00*
7. Sugar	*3*	*1*	*00*	*1*	*80*
8. Salt	*1*	*0*	*25*	*0*	*50*
	100				

SOLUTION. Computation of the weighted index number for the food group for 1969 with 1959 as the base.

Items in the Food Groups	Prices (per kg.)				Price Relatives with 1939 prices taken as 100	Weights	Weighted Price Relative
	1959 (Base year)		1969				
	Rs.	p.	Rs.	p.			
1. Wheat	0	40	1	20	300	40	12,000
2. Rice	0	80	2	00	250	20	5,000
3. Gram	1	00	1	60	160	15	2,400
4. Arhar dal	1	30	1	95	150	5	750
5. Milk	0	60	1	20	200	6	1,200
6. Mustard oil	2	00	4	00	200	10	2,000
7. Sugar	1	00	2	80	180	3	540
8. Salt	0	25	0	50	200	1	200
Total						100	24,090

The weighted index number for the food group for 1969 is equal to $\frac{24,090}{100}$ or 241.

PROBLEM 6.11

Group	*Prices*		*Quantity*	
	Base year	*1975*	*Base year*	*1975*
1	*12*	*20*	*50*	*120*
2	*10*	*12*	*100*	*80*
3	*14*	*15*	*60*	*70*
4	*16*	*18*	*30*	*50*
5	*18*	*20*	*40*	*40*
6	*22*	*15*	*70*	*60*
7	*20*	*16*	*90*	*100*
8	*15*	*18*	*80*	*80*

Find the index number for 1975 by

(1) Base year weighting

(2) Current year weighting

(3) Fisher's Ideal Formula

SOLUTION. Computation of the Index numbers of Prices for 1975

Crops	Base year		Current year 1975		Base year weighting		Current year weighting	
	Prices p_0	Quantity q_0	Prices p_1	Quantity q_1	p_0q_0	p_1q_0	p_0q_1	p_1q_1
1	12	50	20	120	600	1000	1440	2400
2	10	100	12	80	1000	1200	800	960
3	14	60	15	70	840	900	980	1050
4	16	30	18	50	480	540	800	900
5	18	40	20	40	720	800	720	800
6	22	70	15	60	1540	1050	1320	900
7	20	90	16	100	1800	1440	2000	1600
8	15	80	18	80	1200	1440	1200	1440
					Σp_0q_0 =8180	Σp_1q_0 =8370	Σp_0q_1 =9260	Σp_1q_1 = 10005

(1) *Index number of prices by base year weighting*

$$\text{Index Number for 1975}=\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times 100=\frac{8370}{8180}\times 100=102.3$$

(2) *Index number of prices by current year weighting*

$$\text{Index Number for 1975}=\frac{\Sigma p_1q_1}{\Sigma p_0q_1}\times 100=\frac{10050}{9260}\times 100=108{\cdot}5$$

(3) *Index number of prices for 1947 by Fisher's "Ideal" Formula.*

$$\text{Index Number for 1975}=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}}\times 100$$

$$=\sqrt{\frac{8370}{8180}\times\frac{10050}{9260}}\times 100$$

$$=1.053\times 100=105.3$$

Thus, the index number for 1975

(1) by base year weighting is=102.3,
(2) by current year weighting is=108.5, and
(3) by Fisher's Ideal Formula is=105.3.

PROBLEM 6.12 *Given the following data, what index numbers would you use for purposes of comparison? Give reasons.*

	Rice		*Wheat*		*Jowar*	
Year	*Price*	*Quantity*	*Price*	*Quantity*	*Price*	*Quantity*
1967	*9.3*	*100*	*6.4*	*11*	*5.1*	*5*
1974	*4.5*	*90*	*3.7*	*10*	*2.7*	*3*

Prices and quantities are given in arbitrary units.

SOLUTION. Since the statistics of quantities for the base year as well as the current year are given, we would use Fisher's Ideal Index Number for purposes of comparison as it will have the two basic qualities, *viz.*, Time reversibility and Factor reversibility.

Computation of the Fisher's Ideal Index Number for 1974 with 1967 as base.

Commodities	1967 (Base year)		1974					
	Price p_0	Quantity q_0	Price p_1	Quantity q_1	p_0q_0	p_1q_0	p_0q_1	p_1q_1
1. Rice	9.3	100	4.5	90	930.0	450.0	837.0	405.0
2. Wheat	6.4	11	3.7	10	70.4	40.7	64.0	37.0
3. Jowar	5.1	5	2.7	3	25.5	13.5	15.3	8.1
					Σp_0q_0 = 1025.9	Σp_1q_0 = 504.2	Σq_0q_1 = 916.3	Σp_1p_1 = 450.1

Substituting the above value in the Fisher's "Ideal" Formula

$$\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_1} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

We get,

$$\text{Index Number for } 1974 = \sqrt{\frac{504.2}{1025.9} \times \frac{450.1}{916.3}} \times 100$$

$$= .4911 \times 100 = 49.1$$

Thus, the index number for 1974 by Fisher's Ideal Formula with 1967 as the base year is 49.1.

PROBLEM 6.13 *Explain what is meant by the Factor Reversal Test. Construct with the help of the data given below Fisher's Ideal Index, and show how it satisfies the factor reversal test.*

	Estimated total produce in thousand tons in district Saran		*Harvest price per kg. in district Saran*	
	1957—58	*1958—59*	*1957—58*	*1958—59*
			pl	*p.*
Winter Rice	*17*	*26*	*56*	*50*
Barley	*107*	*83*	*32*	*30*
Maize	*62*	*48*	*41*	*28*

SOLUTION. Construction of Index Numbers of Prices by Fisher's Idea Formula.

Commodities	Base year (1957-58)		Current year (1958-59)		p_0q_0	p_1q_1	p_1q_0	p_0q_1
	Price in paisas per kg. (p_0)	Quantity (q_0)	Price in paisas per kg. (p_1)	Quantity (q_1)				
1. Winter Rice	56	17	50	26	952	1300	850	1456
2. Barley	32	107	30	83	3424	2490	3210	2656
3. Maize	41	62	28	48	2542	1344	1736	1968
					Σp_0q_0 $=6918$	Σp_1q_1 $=5134$	Σp_1q_0 $=5796$	Σp_0q_1 $=608$

Substituting the above values in the Fisher's Ideal Formula for Inde numbers,

$$\text{Index Number for current year} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

We get,

$$\text{Index Number for 1958—59} = \sqrt{\frac{5796}{6918} \times \frac{5134}{6080}} \times 100$$

$$= .84 \times 100 = 84.$$

Factor Reversal Test is satisfied if :—

$$P_{01} \times Q_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

Where P_{01} stands for the price change for the current year on the base ye and Q_{01} for the quantity change for the current year on the base year.

Now according to Fisher's Ideal formula.

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \text{ ; and similarly } Q_{01} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_1q_0}}$$

$$\therefore P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \frac{\Sigma p_0q_1}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_1q_0}}$$

$$= \sqrt{\frac{\Sigma p_1q_1}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_0}} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

Thus, the Fisher's Ideal Formula satisfies the Factor Reversal Test.

PROBLEM 6.14 *Prove using the following data that the Factor Revers Test is satisfied by Fisher's Ideal Formula for Index No.*

Commodity	*Base year Price*	*Base year Quantity*	*Current year Price*	*Current year Quantity*
A	*6*	*50*	*10*	*56*
B	*2*	*100*	*2*	*120*
C	*4*	*60*	*6*	*60*
D	*10*	*30*	*12*	*24*
E	*8*	*40*	*12*	*36*

SOLUTION. Calculation of values of Σp_0q_0, Σp_1q_0, Σp_0q_1 and Σp_1q_1.

Commodity	Base year Price p_0	Base year Quantity q_0	Current year Price p_1	Current year Quantity q_1	p_0q_0	p_1q_0	p_0q_1	p_1q_1
1. A	6	50	10	56	300	500	336	560
2. B	2	100	2	120	200	200	240	240
3. C	4	60	6	60	240	360	240	360
4. D	10	30	12	24	300	360	240	288
5. E	8	40	12	36	320	480	288	432
					$\Sigma p_0q_0 = 1360$	$\Sigma p_0q_1 = 1900$	$\Sigma p_0q_1 = 1344$	$\Sigma p_1q_1 = 1880$

Factor Reversal Test is satisfied if :—

$$P_{01} \times Q_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

Where P_{01} stands for the price change for the current year on the base year and Q_{01}, for the quantity change for the current year on the base year.

Now, according to Fisher's Ideal Formula for Index No.

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \text{ and}$$

$$Q_{01} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_0}{\Sigma p_1q_0}}$$

Therefore,

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_0}{\Sigma p_0q_1} \times \frac{\Sigma p_0q_1}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_1q_0}}$$

Substituting the values of Σp_0q_0, Σp_1q_0, Σp_0q_1 and Σp_1q_1 from the above table, we get

$$P_{01} \times Q_{01} = \sqrt{\frac{1900}{1300} \times \frac{1880}{1344} \times \frac{1344}{1360} \times \frac{1880}{1900}}$$

$$= \sqrt{\frac{1880}{1360} \times \frac{1880}{1360}} = \frac{1880}{1360}$$

Now $\frac{\Sigma p_1q_1}{\Sigma p_0q_0}$ is also equal to $\frac{1880}{1360}$

Thus, it is proved that the Factor Reversal Test is satisfied by Fisher's Ideal Formula.

PROBLEM 6.15 *Construct with the help of the data given below Fisher's Ideal Index, and show how it satisfies the Time Reversal Test.*

Commodity	*Unit*	*Base year Price Rs.*	*Base year Quantity*	*Current year Price*	*Current year Quantity*
Wheat	*p. 40 kg.*	8	50	20	60
Ghee	*p. kg.*	2	15	6	10
Firewood	*p. 40 kg.*	1	20	2	25
Sugar	*p. five kg.*	2	10	5	8
Cloth	*p. mt.*	1	40	3	30

SOLUTION. Computation of the Fisher's Index Number for the current year.

Commodity	Base year		Current year		p_0q_0	p_1q_0	p_0q_1	p_1q_1
	Price p_0	Quantity q_0	Price p_1	Quantity q_1				
1. Wheat	8	50	20	60	400	1000	480	1200
2. Ghee	2	15	6	10	30	90	20	60
3. Firewood	1	20	2	25	20	40	25	50
4. Sugar	2	10	5	8	20	50	16	40
5. Cloth	1	40	3	30	40	120	30	90
					Σp_0q_0 $=510$	Σp_1q_0 $=1300$	Σp_0q_1 $=571$	Σp_1q_1 $=1440$

Substituting the above values in the Fisher's Ideal Formula for Index No.

$$\text{Index No. for current year} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

we get,

$$\text{Index No. for current year} = \sqrt{\frac{1300}{510} \times \frac{1440}{571}} \times 100$$

$$= 2{\cdot}535 \times 100 = 253.5$$

Thus, the Fisher's Ideal Index Number for current year is 253.5

Time reversal Test is satisfied if

$P_{01} \times P_{10} = 1$

Where P_{01} stands for the price change for the current year on the base year, and P_{10} for the price change for the base year on the current year.

According to the Fisher's Ideal Formula

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}}\text{, and similarly } P_{10} = \sqrt{\frac{\Sigma p_0q_0}{\Sigma p_1q_0} \times \frac{\Sigma p_0q_1}{\Sigma p_1q_1}}$$

Multiplying p_{01} and p_{10}, we get

$p_{01} \times p_{10} = \sqrt{1} = 1$

Thus, the Fisher's Ideal Formula satisfies the time reversal test.

There is yet another way of looking at the reversibility of index numbers. If a relative shows an average increase of say, 25 per cent from the base year to the current year then this should also be capable of being described as a decrease of 20 per cent from the current year to the base year. From the given data we find,

(*a*) Index Number for the current year on the base year, as culculated is 253.5.

(*b*) and index number for the base year on the current year

$$= \sqrt{\frac{\Sigma p_0q_0}{\Sigma p_1q_0} \times \frac{\Sigma p_0q_1}{\Sigma p_1q_1}} \times 100$$

$$= \sqrt{\frac{510}{1300} \times \frac{571}{1440}} \times 100 = .3945 \times 100 = 39.45$$

From the above we find that the level of prices in the current year is 153.5 per cent higher than the prices of the base year. If this is also capable of being described as a decrease of 60.55 per cent from the current year to the base year, the time reversal test is satisfied. Thus, if the index number for the current year on the base year is 253.5, it should be (100—60.55) = 39.45 for the base year on the current year, which actually is the case. Therefore the Fisher's Ideal Index can be said to satisfy the time reversal test.

PROBLEM 6.16 *The following figures show the imports of cotton piecegoods into India from Great Britain during 1953—54 and a few later years. Find (a) index numbers of quantity (b) index numbers of value, and (c) index numbers of price, using the figures of 1953 —54 as base.*

Year	*Quantity million metres*			*Value million £*		
	Grey (unbleached)	*White (bleached)*	*Coloured (Printed or dyed)*	*Grey (unbleached)*	*White (bleached)*	*Coloured (Printed or dyed)*
1953—54	*1534*	*793*	*832*	*17.0*	*9.5*	*11.9*
1969—70	*926*	*474*	*483*	*15.7*	*10.0*	*11.4*
1970—71	*365*	*272*	*246*	*5.2*	*4.7*	*5.1*
1971—72	*249*	*280*	*223*	*2.9*	*4.0*	*3.8*

Ignore the effects of exchange variation between the countries.

SOLUTION. (*a*) Computation of the Index Numbers of Quantity.

Cotton goods	Actual quantity in million yards				Percentage with base 1953—54 = 100			
	1953—54	1969—70	1970—71	1971—72	1953—54	1969—70	1970—71	1971—72
1. Grey (unbleached)	1534	926	365	249	100	60	24	16
2. White (bleached)	793	474	272	280	100	60	34	35
3. Coloured (printed or dyed)	832	483	246	223	100	58	30	27
Index numbers as geometric mean of relatives					100	59.3	29.0	24.7

(*b*) Computation of the Index Numbers of Value.

Cotton goods	Actual value in million £				Percentage with base 1953—54 = 100			
	1953	1969	1970	1971	1973	1969	1970	1971
1. Grey (unbleached)	17	15.7	5.2	2.9	100	92	31	17
2. White (bleached)	9.5	10.0	4.7	4.0	100	105	49	42
3. Coloured (Printed or dyed)	11.9	11.4	5.1	3.8	100	96	43	32
Index numbers as geometric mean of relatives					100	97.3	40.3	28.4

(*c*) Index Number of Price.

To find the index numbers of prices, goods are valued in all the years separately at their prices in the base year. The ratio of the value actually recorded to the value thus found out multiplied with 100 is the index number of the year selected.

Cotton goods	Value actually recorded				Value at the price in the base year			
	1953 —54	1969 —70	1970 —71	1971 —72	1953 —54	1969 —70	1970 —71	1971 —72
1. Grey 2. White 3. Coloured	1.7 9.5 11.9	15·7 10.0 11.4	5.2 4.7 5.1	2.2 4.0 3·8	17 9.5 11.9	10.3 5.7 6.9	4.0 3.3 3.5	2.8 3.3 3.2
Total	38.4	37.1	15.0	10.7	38.4	22.9	10.8	9.3
Index numbers of prices as ratios of total values actually recorded to total values at base year prices (expressed as percentages)					100	162	139	115

PROBLEM 6.17. *The following are the group index numbers and the group weights of an average working class family's budget. Construct the cost of living index number by assigning the given weights.*

Groups	*Index No.*	*Weight*
Food	*352*	*48*
Fuel and Lighting	*220*	*10*
Clothing	*230*	*8*
Rent	*160*	*12*
Miscellaneous	*190*	*15*

SOLUTION. Computation of the cost of living *index* number.

Group	Index No.	Weights	Weighted Relatives
1. Food	352	40	16896
2. Fuel and Lighting	220	10	2200
3. Clothing	230	8	1840
4. Rent	160	12	1920
5. Miscellaneous	190	15	2850
	Total	93	25706

$$\text{Cost of living index number} = \frac{\Sigma\text{weighted relatives}}{\Sigma\text{weights}} = \frac{25706}{93}$$

$$= 276.4$$

PROBLEM 6.18. *Calculate a cost of living index from the following indices the weights being Food 60, Fuel and Light 8, Clothing 12, Rent 16, and Miscellaneous 4, Total 100.*

	(i) *Food*	*(ii)* *Rent*	*(iii)* *Clothing*	*(iv)* *Fuel and Light*	*(v)* *Miscellaneous*
1964	*100*	*100*	*100*	*100*	*100*
1965	*102*	*100*	*103*	*100*	*97*
1966	*106*	*102*	*105*	*101*	*98*
1967	*104*	*103*	*106*	*102*	*99*
1968	*107*	*105*	*108*	*101*	*102*

SOLUTION. Calculation of the cost of living index numbers

Items	1965			1966		1967		1968	
	Weights	Index No.	Weighted Relatives	Index No.	Weighted Relatives	Index No.	Weighted Relatives	Index No.	Weighted Relatives
1. Food	60	102	6120	106	6360	104	6240	107	6420
2. Fuel and Light	8	100	800	101	808	102	816	101	808
3. Clothing	12	103	1236	105	1260	106	1272	108	1296
4. Rent	16	100	1600	102	1632	103	1648	105	1680
5. Miscellaneous	4	97	388	98	392	99	396	102	408
Total	100		10144		10452		10372		10612
Cost of living Index No.	$\frac{10144}{100} = 101.44$			$\frac{10452}{100} = 104.52$		$\frac{10372}{100} = 103.72$		$\frac{10612}{100} = 106.12$	

Thus, the cost of living index number for 1965, 1966, 1967 and 1968 based on 1964 are 101.44, 104.52. 103.72 and 106.12 respectively.

PROBLEM 6.19 *From the following information prepare the Index Number for the Miscellaneous Group taken for the construction of the Allahabad working class cost of living Index Number.*

Miscellaneous Group

No.	Articles	Unit of quantity	Weight assigned	Average Prices in 1969 Rs.	p.	Prices in 1971 Rs.	p.
1.	Barber	per shave	13	1	00	1	50
2.	Soap	per bar	9	0	60	0	80
3.	Medicine	per bottle	3	2	00	4	00
4.	Supari	per kg.	25	6	00	8	00
5.	Bidis	per bundle	22	0	30	0	40
6.	Travelling Expenses		27	2	60	3	00
7.	Newspaper	per copy	1	0	20	0	25

SOLUTION. Computation of the Index Number for the miscellaneous group.

No.	Articles	Unit of quantity	Average price in 1969 (Base year) Rs.	p.	Price in 1971 Rs.	p.	Price Relative with 1969 = 100	Weight	Weighted Relatives
1.	Barber	per shave	1	00	1	50	150	13	1950
2.	Soap	per bar	0	60	0	80	133	9	1197
3.	Medicine	per bottle	2	00	4	00	200	3	600
4.	Supari	per kg.	6	00	8	00	133	25	3325
5.	Bidis	per bundle	0	30	0	40	133	22	2926
6.	Travelling Expenses	..	2	60	3	00	150	27	4050
7.	Newspaper	per copy	0	20	0	25	125	1	125
							Total	100	14,173

The index number for the miscellaneous group is $\frac{14173}{100} = 141.73$

PROBLEM 6.20. *Construct the cost of living index number for 1950 on the basis of 1946 from the following data using the family Budget Method.*

Articles	*Quantity consumed in 1946*	*Unit*	*Price in 1946 Rs.*	*Price in 1950 Rs.*
Rice	*2 Quintal*	*per 40 kg*	*12*	*16*
Bajra	*2 ,,*	*per 40 kg.*	*8*	*10*
Wheat	*40 kg.*	*per 40 kg.*	*10*	*20*
Gram	*40 ,,*	*per 40 kg.*	*6*	*12*
Arhar	*2 Quintal*	*per 40 kg.*	*8*	*12*
Other pulses	*80 kg.*	*per 40 kg.*	*6*	*8*
Ghee	*4 kg.*	*per kg.*	*2.5*	*4*
Gur	*80 kg.*	*per 40 kg.*	*5*	*10*
Salt	*12.5 kg.*	*per 40 kg.*	*8*	*10*
Oil	*24 kg.*	*per 40 kg.*	*40*	*50*
Clothing	*40 mt.*	*per mt.*	*.5*	*1*
Firewood	*2.50 Quintal*	*per 25 kg.*	*1*	*1 6*
Kerosene	*1 tin*	*per tin*	*4*	*7*
House-rent	*..*	*per house*	*24*	*30*

SOLUTION. Construction of cost of Living Index Number.

Articles	Quantities consumed in base year	Unit	Price in base year (1946)	Price in current year (1950)	Price Relatives for current year	Values consumed in base year (weights)	Product of price relatives and weights
				Rs.		Rs.	
Rice	200 kg.	per 40 kg.	12	16	133.3	60	7998
Bajra	200 ,,	,, ,, ,,	8	10	125	40	5000
Wheat	40 ,,	,, ,, ,,	10	20	200	10	2000
Gram	40 ,,	,, ,, ,,	6	12	200	6	1200
Arhar	200 ,,	,, ,, ,,	8	12	150	40	6000
Other pulses	80 ,,	,, ,, ,,	6	8	133.3	12	1599.6
Ghee	4 ,,	,, ,,	2.5	4	160	10	1600 0
Gur	80 ,,	,, ,, ,,	5	10	200	10	2000 0
Salt	12.5 ,,	,, ,, ,,	8	10	125	2.5	312.5
Oil	24 ,,	,, ,, ,,	40	50	125	24	3000
Clothing	40 mt.	,, ,, ,,	.5	1	200	20	4000
Firewood	250 kg.	,, 25 ,,	1	1.6	160	10	1600
Kerosene	1 tin	per tin	4	7	175	4	700
House rent	..	per house	24	30	125	24	3000
						272.5	40010.1

$$\text{Index Number for 1950} = \frac{\Sigma(\text{price relative} \times \text{weight})}{\Sigma \text{weight}} = \frac{40010.1}{272.5} = 146.8$$

Thus, the cost of living index number for 1950 is 146.8 or 147 Approx.

PROBLEM 6.21. *An enquiry into the budgets of the middle class families in a city in England gave the following information :*

Expenses on	*Food*	*Rent*	*Clothing*	*Fuel*	*Misc.*
	35%	*15%*	*20%*	*10%*	*20%*
Prices (1968)	*£150*	*£30*	*£75*	*£25*	*£40*
Prices (1969)	*£145*	*£30*	*£65*	*£23*	*£45*

What changes in cost of living figures of 1969 as compared with that of 1968 are seen?

SOLUTION. Computation of cost of living index number for 1969 with 1968 as the base.

Items of Expenditure	1968		1969		Weight	Weighted Relative
	Price	1968 prices taken as 100	Prices	Price relatives based on 1968		
	£		£			
1. Food	150	100	145	97	35	3395
2. Rent	30	100	30	100	15	1500
3. Clothing	75	100	65	87	20	1740
4. Fuel	25	100	23	92	10	920
5. Misc.	40	100	45	112	20	2240
	Total				100	9795

Cost of living Index Number for 1969 based on $1968 = \frac{9795}{100} = 97.95$

Thus a fall of 2.05% has taken place in the cost of living of middle class families in the given city of England in 1969 as compared with 1968.

PROBLEM 6.22. *The following are the index numbers of prices (Base 1959 = 100).*

Year	*Index*
1959	*100*
1960	*110*
1961	*120*
1962	*200*
1963	*320*
1964	*400*
1965	*410*
1966	*400*
1967	*380*
1968	*370*
1969	*350*
1970	*360*
1971	*348*
1972	*340*
1973	*320*
1974	*300*

Shift the base from 1959 to 1966 and recast the index numbers.

SOLUTION

Shifting of base from 1959 to 1966

Year	Index Nos. 1959 Base	Index Nos. 1966 Base
1959	100	$100 \times \frac{100}{400} = 25$
1960	110	$110 \times \frac{100}{400} = 27.5$
1961	120	$120 \times \frac{100}{400} = 30$
1962	200	$200 \times \frac{100}{400} = 50$
1963	320	$320 \times \frac{100}{400} = 80$
1964	400	$400 \times \frac{100}{400} = 100$
1965	410	$410 \times \frac{100}{400} = 102.5$
1966	400	$400 \times \frac{100}{400} = 100$
1967	380	$380 \times \frac{100}{400} = 95$
1968	370	$370 \times \frac{100}{400} = 92.5$
1969	350	$350 \times \frac{100}{400} = 87.5$
1970	360	$360 \times \frac{100}{400} = 90$
1971	348	$348 \times \frac{100}{400} = 87$
1972	340	$340 \times \frac{100}{400} = 85$
1973	320	$320 \times \frac{100}{400} = 80$
1974	300	$300 \times \frac{100}{400} = 75$

PROBLEM 6.23 *Given below are two sets of indices, one with 1959 as* *and the other with 1967 as base.*

(a) Year	*Index Nos.*	*(b) Year*	*Index Nos.*
1959	*100*	*1967*	*100*
1960	*120*	*1968*	*110*
1961	*130*	*1969*	*90*
1962	*200*	*1970*	*98*
1963	*300*	*1971*	*101*
1964	*350*	*1972*	*110*
1965	*370*	*1973*	*98*
1966	*380*	*1974*	*96*
1967	*400*		

The index number (a) with 1959 base was discontinued in 1967. It is desired splice the second index number (b) with 1967 base to the first index number for sake of continuity. How will it be done so that the combined series has a com- n base of 1959.

SOLUTION.

Splicing the Index Number (*b*) with (*a*)

Year	Index No. (*b*) 1959 base	Index No. (*b*) 1967 base	Index No. (*b*) spliced to (*a*) 1959 base
1959	100		
1960	120		
1961	130		
1962	200		
1963	300		
1964	350		
1965	370		
1966	380		
1967	400	100	$100\times\frac{400}{100}=400$
1968		110	$110\times\frac{400}{100}=440$
1969		90	$90\times\frac{400}{100}=360$
1970		98	$98\times\frac{400}{100}=392$
1971		101	$101\times\frac{400}{100}=404$
1972		110	$110\times\frac{400}{100}=440$
1973		98	$98\times\frac{400}{100}=392$
1974		96	$96\times\frac{400}{100}=384$

PROBLEM 6.24. *The following table gives the per capita income and cost of living index number of a particular community. Deflate the per capita income by aking into account the rise in the cost of living.*

Year	*Per Capita Income*	*Cost of living Index Numbers Base 1959*
1959	*65*	*100*
1960	*70*	*110*
1961	*75*	*120*
1962	*80*	*130*
1963	*90*	*150*
1964	*100*	*200*
1965	*120*	*250*
1966	*150*	*350*

SOLUTION. Deflating per Capita Income.

Year	Per Capita Income in Rs.	Cost of living Index Base 1959	Deflated or Real pe Capita Income in Rs.
1959	65	100	65.0
1960	70	110	$70\times\frac{100}{110}=63.6$
1961	75	120	$75\times\frac{100}{120}=62.5$
1962	80	130	$80\times\frac{100}{130}=61.5$
1963	90	150	$90\times\frac{100}{150}=60.0$
1964	100	200	$100\times\frac{100}{200}=50.0$
1965	120	250	$120\times\frac{100}{250}=48.0$
1966	150	350	$150\times\frac{100}{350}=42.9$

REVISIONAL PROBLEMS-6

6.25 The prices of four Commodities in a place during two years are below. Compute the four simple index numbers with the given da 1966 with 1965 as base year.

Year	Price in rupees per Unit for Commodities I	II	III	IV
1965	156	152	156	154
1966	307	310	315	292

[*Commodity*	*I*	*II*	*III*
Index No.	196.8	203.9	109.9

6.26 Find index numbers for the year 1967, 1968 & 1969 by the chain method, with base year 1966, from the following table :

Year	1966	1967	1968	1969
Link Index	100	110	95.5	109.5

(*Chain Indices* : *110, 105.5 and 115.0*

6.27 Calculate the cost of living Index Number by using (*i*) the weighted metic mean (*ii*) the weighted geometric mean :

Group	Index Number	Weights
1. Food	350	5
2. Fuel and Lighting	200	1
3. Clothing	240	1
4. House Rent	160	1
5. Miscellaneous	250	2

(*Index nos. based on mean=285*
Index nos. based on G. M.=275

6.28 In the preparation of rural index number in a certain centre the follo results were obtained :

Group	Wt.	Group Oct. 1966	Index number Nov., 1966
Food	81	323	344
Lighting	2	190	186
Clothing	8	432	397
Misc.	9	369	377

Find the index number of rural prices for the two months.

(Index No. for Oct.,=333.2; Index No. for Nov.,=348.5)

29 Compute Laspeyre's, Paasche's and thence Fisher's index for the following data.

		Wheat	Rice	Maize
Quantity	1959	15	5	10
	1964	12	4	5
Price Rs.	1959	15	20	4
	1964	22	27	7

(Ans: Po,=146.6)

30 Given the following data, what index number will you use for purposes of comparison? Give reasons.

	Rice		Wheat		Jowar	
Year	Price	Qty	Price	Qty.	Price	Qty
1957	9.3	104	6.4	11	5.1	5
1967	4.5	90	3.7	10	2.7	3

(Po_1=49.1, Fisher's Index)

31 In calculating cost of living index the following weights were used :

Food $8\frac{1}{2}$, Rent 2, Clothing $2\frac{1}{2}$, Fuel and Light 1, Miscellaneous $1\frac{1}{2}$. Calculate the Index number for a data when the percentage increase in prices of the various items over prices of July, 1968=100 were 31, 57, 90, 75 and 88 respectively.

(Cost of Living Index=152.2)

32 From the Chain base index numbers given below prepare fixed base index numbers and verify the calculations:

Year	1962	1963	1964	1965	1966
Index	80	110	120	105	125

(Ans. 80, 88, 105.6, 110.9 and 105.33)

33 From the figures given below calculate the index numbers of real wages taking 1968 as base:

Year	Average wages	Index No. 1965=100
1965	120	100
1966	132	120
1967	143	130
1968	150	150
1969	171	180

(Ans: 120, 110, 110, 100, 95)

6.34 In 1950, a Statistical Bureau started an index of production based on 1944, with the following results:

Year	1944 (Base)	1950	1959
Index	100	120	200

In 1960, the Bureau reconstructed the index on a plan with base 1959, 1960,

Year	1959 (Base)	1965
Index	100	150

In 1966, the Bureau again reconstructed the index on yet another plan with the base year 1965.

Year	1965 (Base)	1969
Index	100	120

Obtain a continuous series with base 1965, by splicing the three series

(Ans :	*1944*	*1950*	*1959*	*1965*	*1969)*
	33	*40*	*67*	*100*	*120*

Analysis of Time Series

In this chapter questions on the analysis of time series have been solved. S questions deal with the separation of the trend from the fluctuations by mea the Method of Moving Averages, and the Method of Least Squares. Ques have also been solved on the analysis of the regular and irregular fluctuations time series. Calculation of the Seasonal Variations Index has also been show some questions.

PROBLEM 7.1. *Explain how you will deal with a time series and illus your remarks with the help of the following series of annual figures for the pe 1941 – 1970.*

Period	*Annual Values*
1941—1950	*208, 223, 225, 222, 239, 242, 238, 252, 257,*
1951—1960	*273, 270, 268, 288, 284, 282, 300, 303, 298,*
1961—1970	*317, 309, 329, 333, 327, 345, 344, 343, 362,*

SOLUTION. Computation of the trend, and the short-time fluctuation the annual values for the period 1941—1970, (assuming a five-yearly cycle).

Year	Annual values	Add in 5's	5-yearly moving average (Trend)	Short-time Fluctuatio
1941	208			
1942	223			
1943	225	1117	223	+ 2
1944	222	1151	230	– 8
1945	239	1166	233	+ 6
1946	242	1193	239	+ 3
1947	238	1228	246	– 8
1948	252	1239	248	+ 4
1949	257	1270	254	+ 3
1950	250	1302	260	–10
1951	273	1318	264	+9
1952	270	1349	270	0
1953	268	1383	277	– 9
1954	288	1392	278	+10
1955	284	1422	284	0
1956	282	1457	291	– 9
1957	300	1467	299	+ 7
1958	303	1496	293	+ 4
1959	298	1531	306	– 8

1960	313	1540	308	+ 5
1961	317	1566	313	+ 4
1962	309	1601	320	—11
1963	329	1615	323	+ 6
1964	333	1643	329	+ 4
1965	327	1678	336	− 9
1966	345	1692	338	+ 7
1967	344	1721	344	0
1968	343	1754	351	— 8
1969	362			
1970	360			

PROBLEM 7.2. *Explain the use of moving averages in the analysis of time series. Find out an approximate moving average for the following series :*

1951	*506*	*1958*	*738*	*1965*	*1276*
1952	*620*	*1959*	*663*	*1966*	*898*
1953	*1036*	*1960*	*777*	*1967*	*814*
1954	*673*	*1961*	*1189*	*1968*	*929*
1955	*588*	*1962*	*188*	*1969*	*1360*
1956	*696*	*1963*	*745*	*1970*	*961*
1957	*1116*	*1964*	*854*	*1971*	*926*

SOLUTION. There is a four-yearly cycle in the given series as the time distance between the consecutive peaks which lie in 1953, 1957, 1961, 1965, 1969, is 4 years.

Computation of 4-yearly moving average.

Year	Annual values	Add in fours	Add in pairs	(Divided by 8) 4-yearly moving average (Trend)
1951	506			
1952	620			
1953	1036	2835	5752	719
1954	673	2917	5910	739
1955	588	2993	6066	758
1956	696	3073	6211	776
1957	1116	3138	6351	794
1958	738	3213	6507	813
1959	663	3294	6661	833
1960	777	3367	6814	852
1961	1189	3447	6976	872
1962	818	3529	7126	891
1963	745	3597	7281	910
1964	845	3684	7448	931
1965	1276	3764	7597	950
1966	898	3833	7750	969
1967	814	3917	7918	990
1968	929	4001	8065	1008
1969	1360	4064	8240	1030
1970	961	4176		
1971	926			

PROBLEM 7.3. *The following are the figures for the infantile mortality rate in England and Wales (deaths of infants under one year of age per 1.000 live births).*

Year	*Rate*	*Year*	*Rate*	*Year*	*Rate*
1951	*77*	*1959*	*60*	*1967*	*53*
1952	*69*	*1960*	*66*	*1968*	*51*
1953	*75*	*1961*	*65*	*1969*	*57*
1954	*75*	*1962*	*64*	*1970*	*60*
1955	*70*	*1963*	*59*	*1971*	*61*
1956	*70*	*1964*	*57*	*1972*	*49*
1957	*65*	*1965*	*59*	*1973*	*45*
1958	*74*	*1966*	*58*	*1974*	*46*
				1975	*43*

Fit a simple moving average of fives to the series and apply a further simpl moving average of five to the result.

SOLUTION. Computation of a simple moving average of fives of the infantil mortality rate in England and Wales and a further simple moving average of fives o the result.

Year	Rate (Deaths of infants under one year of age per 1000 live births)	Add in fives	5-yearly moving average of the rates	Add in fives	Further simple moving average of fives of the result
1951	77				
1952	69				
1953	75	366	73		
1954	75	359	72		
1955	70	355	71	355	71
1956	70	354	71	349	70
1957	65	339	68	343	69
1958	74	335	67	338	68
1959	60	330	66	330	66
1960	66	329	66	324	65
1961	65	314	63	318	64
1962	64	311	62	311	62
1963	59	304	61	302	60
1964	57	297	59	295	59
1965	59	286	57	289	58
1966	58	278	56	284	57
1967	53	278	56	281	56
1968	51	279	56	280	56
1969	57	282	56	278	56
1970	60	278	56	271	55
1971	61	272	54	267	53
1972	49	261	52		
1973	45	244	49		
1974	46				
1975	43				

PROBLEM 7.4. *The following table gives the Bank Clearings in the Kanpur city for the years 1946 to 1970 in millions of rupees. Find the trend.*

1946	*52.7*	*1954*	*79.3*	*1962*	*159.6*
1947	*79.4*	*1955*	*103.6*	*1963*	*177.4*
1948	*76.3*	*1956*	*97.3*	*1964*	*178.6*
1949	*66.0*	*1957*	*92.1*	*1965*	*235.8*
1950	*68.6*	*1958*	*100.7*	*1966*	*243.2*
1951	*93.8*	*1959*	*94.6*	*1967*	*194.4*
1952	*104.7*	*1960*	*83.0*	*1968*	*217.9*
1953	*87.2*	*1961*	*110.6*	*1969*	*214.0*
				1970	*256.7*

SOLUTION. Computation of the trend of the bank clearings in the Bombay city for the years 1946 to 1970.

(1) Year	(2) Bank clearings	(3) Add in nines	(4) Trend (9 yearly moving average)	(1) Year	(2) Bank clearings	(3) Add in nines	(4) Trend (9 yearly moving average)
1946	52.7			1959	94.6	1019.2	113.2
1947	79.4			1960	83.0	1094.2	121.6
1948	76.3			1961	110.6	1232.7	137.0
1949	66.0			1962	159.6	1383.5	153.7
1950	68.6	708.0	78.7	1963	177.4	1473.2	164.1
1951	93.8	758.9	84.3	1964	178.6	1600.5	177.8
1952	104.7	776.8	86.3	1965	235.8	1731.5	192.4
1953	87.2	792.9	88.1	1966	243.2	1877.6	208.6
1954	79.3	827.6	92.0	1967	194.4		
1955	103.6	853.6	94.8	1968	217.9		
1956	97.3	842.8	93.6	1969	214.0		
1957	92.4	848.7	94.3	1970	256.7		
1958	100.7	921.1	102.3				

PROBLEM 7.5. *Assuming a ten-yearly cycle for the following series relating to Index Numbers of the Retail Price of wheat in India (1939=100), give the trend values, and represent graphically the shorttime fluctuations with the trend removed.*

Year	*Annual Average*	*Year*	*Annual Average*	*Year*	*Annual Average*
1946	*155*	*1954*	*200*	*1962*	*315*
1947	*168*	*1955*	*227*	*1963*	*356*
1948	*226*	*1956*	*193*	*1964*	*246*
1949	*203*	*1957*	*205*	*1965*	*294*
1950	*170*	*1958*	*270*	*1966*	*281*
1951	*153*	*1959*	*341*	*1967*	*267*
1952	*170*	*1960*	*310*	*1968*	*264*
1953	*177*	*1961*	*360*	*1969*	*262*

SOLUTION. Computation of the trend of Index Numbers of the Retail Price of Wheat in India.

Year (1)	Annual Average (2)	Add in tens (3)	Add in pairs (4)	(Divide by 20) 10-yearly moving average Trend (5)	Short-time fluctuations (6)
1946	155				
1947	168				
1948	226				
1949	203				
1950	170				
1951	153	1849	3736	186.8	—33.8
1952	170	1887	3811	190.5	—20.5

1953	177	1924	3892	194.6	—17.6
1954	200	1968	4074	203.7	— 3.7
1955	227	2106	4352	217.6	+ 9.4
1956	193	2246	4699	234.9	—41.9
1957	205	2453	5051	252.5	—47.5
1958	270	2598	5375	268.7	+ 1.3
1959	341	2777	5600	280.0	+61.0
1960	310	2823	5713	285.7	+24.3
1961	360	2890	5868	293.4	+66.6
1962	315	2978	6018	500.9	+14.1
1963	356	3046	6074	303.7	+52.3
1964	246	3034	5989	299.4	—53.4
1965	294	2955			
1966	281				
1967	267				
1968	264				
1969	262				

PROBLEM 7.6. *Below are given figures of production (in thousand tons) of sugar factory.*

Year	*Production in thousand tons*
1966	*80*
1967	*90*
1968	*92*
1969	*83*
1970	*94*
1971	*99*
1972	*92*

(*a*) *Find the slope of a straight line trend to these figures.*

(*b*) *Plot these figures on a graph and show the trend line.*

(*c*) *Do these figures show a rising trend or falling trend ? How do you arrive at your conclusion ?*

SOLUTION. Computation of trend of the production of a sugar factory by the Method of Least Squares.

Year	Production (in thousand tons) (*m*)	Time-Deviation from the middle year (1969) (*d*)	Square of deviation $(d)^2$	Product of Production and time deviation *dm*	Trend ordinates (thousand tons.)
1966	80	—3	9	—240	84
1967	90	—2	4	—180	86
1968	92	—1	1	— 92	88
1969	83	0	0	0	90
1970	94	+1	1	+ 94	92
1971	99	+2	4	+198	94
1972	92	+3	9	+276	96
$n=7$	$\Sigma m=630$		$\Sigma d^2=28$	$\Sigma dm=56$	

Average production $= \frac{\Sigma m}{n} = \frac{630}{7} = 90$ thousand tons

Rate of growth $= \frac{\Sigma dm}{\Sigma d^2} = \frac{56}{28} = 2$ thousand tons of sugar per year.

PROBLEM 7.7. *Compute the trend of the Sterling Assets of the Reserve Bank f India by the method of Least Squares.*

Year	*Sterling Assets (crores of Rs.)*
1966—67	*83*
1967—68	*92*
1968—69	*71*
1969—70	*90*
1970—71	*169*
1971—72	*191*

SOLUTION. Computation of trend of the Sterling Assets of the Reserve Bank of India by the Method of Least Squares.

Year	Sterling assets (crores of Rs.) (m)	deviation from the middle year (d)	Deviation squared up (d^2)	Product of sterling assets and deviation (dm)	Trend ordinate (crores of rupees)	
					Calculation	Trend values
1966—67	83	—2.5	6.25	—207.5	(116—11.3 —22.6—22.6)	59.5
1967—68	92	—1.5	2.25	—138.0	(116—11.3 —22.6)	82.1
1968—69	71	— .5	.25	— 35.5	(116—11.3)	104.7
1969—70	90	+ .5	.25	+ 45.0	(116+11.3)	127.3
970—71	169	+1.5	2.25	+253.5	(116+11.3 +22.6)	149.9
1971—72	191	+2.5	6.25	+477.5	(116+11.3 +22.6+22.6)	172.5
$n=6$	$\Sigma m=696$		$\Sigma d^2=$ 17.50	$\Sigma dm=$ +395		

Arithmetic average of the sterling assets $= \frac{\Sigma m}{n} = \frac{696}{6} = 116$ crores of pees.

So 116 crores of rupees is the value of the sterling assets for the middle year.

Rate of growth $= \frac{\Sigma dm}{\Sigma d^2} = \frac{395}{17.5} = 22.6$ crores of rupees per year.

PROBLEM 7.8. *The following are the Monthly Index Numbers of Co Group (Food and Tobacco) issued by the Economic Adviser to the Govt. of Ind*

Prices for week ending 19th August, 1959=100.

Month	*Index*	*Month*	*Index*
1971		*1972*	
Oct.	*127.4*	*July*	*155.8*
Nov.	*127.9*	*August*	*158.9*
Dec.	*127.5*	*September*	*161.0*
1972		*Oct.*	*167.2*
Jan.	*128.4*	*Nov.*	*172.4*
Feb.	*132.3*	*Dec.*	*178.5*
March	*130.5*	*1973*	
Apr.	*136.5*	*Jan.*	*190.8*
May	*144.7*	*Feb.*	*270.0*
June	*152.3*		

Fit a straight line trend to the above data by the method or the Least Sq and exhibit the graph of the data with the trend.

SOLUTION. Computation of the trend of the Monthly Index Numbe commodity group issued by the Economic Adviser to the Govt. of India b method of least squares.

Months	Index with base week ending 19th August 1959=100 (*m*)	Time deviation from the middle month (June 1972) (*d*)	Square of deviation (d^2)	Product of Index and time deviation (*dm*)	Trend or nate
1971					
Oct.	127.4	—8	64	—1019.2	109.4
Nov.	127.9	—7	49	— 895.3	115.3
Dec.	127.5	—6	36	— 765.0	121.2
1972					
Jan.	128.4	—5	25	— 642.0	127.1
Feb.	132.3	—4	16	— 529.2	133.0
March	130.5	—3	9	— 391.5	138.9
April	136.5	—2	4	— 273.0	144.8
May	144.7	—1	1	— 144.7	150.7
June	152.3	0	0	0	156.6
July	155.8	+1	1	+ 155.8	162.5
August	158.9	+2	4	+ 317.8	168.4
Sept.	161.0	+3	9	+ 483.0	174.3
Oct.	167.2	+4	16	+ 668.8	180.2
Nov.	172.4	+5	25	+ 862.0	186.1
Dec.	178.5	+6	36	+1071.0	192.0
1973					
Jan.	190.8	+7	49	+1335.6	197.9
Feb.	270.0	+8	64	+2160.0	203.8
$n=17$	$\Sigma m=2662.1$		$\Sigma d^2=$ 408	$\Sigma dm=$ 2394.1	

Average of the Monthly Index Numbers$=\frac{\Sigma m}{n}=\frac{2662.1}{17}=156.6$

Rate of growth$=\frac{\Sigma dm}{\Sigma d^2}=\frac{2394.8}{408}=5.9.$

PROBLEM 7.9. *Fit a straight line trend by the method of least squares to the llowing data :—*

x	3	5	7	9
y	5	9	12	8

SOLUTION. Fitting of straight line trend by the method of least squares. $=a+bx$.

x	x^2	y	Σxy	Computed values of y $y=0.6x+2.5$
3	9	5	15	4.3
5	25	9	45	5.5
7	49	12	84	6.7
9	81	8	72	7.9
24	164	34	216	24.4

The two normal equations are

$$\Sigma(y)=na+b\Sigma(x) \text{ and}$$
$$\Sigma(xy)=a\Sigma(x)+b\Sigma(x^2)$$

Substituting the values we get,

$$34=4a+24b \quad . \; . \; . \; . \; . \; . \quad (i)$$
$$216=24a+164b. \; . \; . \; . \; . \; . \quad (ii)$$

Solving the equations we get,

$$b=.6$$
$$a=2.5$$

The required equation is $y=.6x+2.5$

Thus if $x=3y=(.6\times 3)+2.5=4.3$

Thus if $x=5y=(.6\times 5)+2.5=5.5$

Thus if $x=7y=(.6\times 7)+2.5=6.7$

Thus if $x=9y=(.6\times 9)+2.5=7.9$

PROBLEM 7.10. *Fit a straight line trend by the method of least squares.*

Year	*World production of Gold (Crore ounces)*
1965	*12.7*
1966	*10.1*
1967	*13.0*
1968	*13.2*
1969	*12.6*
1970	*14.2*
1971	*13.7*
	89.5

SOLUTION.

Fitting of a straight line trend ($y=a+bx$)

Year	Production in Crores (y)	x	xy	x^2	Computed $y=11.243+$
1965	12.7	1	12.7	1	11.629
1966	10.1	2	20.2	4	12.015
1967	13.0	3	39.0	9	12.401
1968	13.2	4	52.8	16	12.787
1969	12.6	5	63.0	25	13.173
1970	14.2	6	85.2	36	13.559
1971	13.7	7	95.9	49	13.945
Total	89.5	28	368.8	140	

Substituting the values in the two normal equations

$\Sigma\,(y)=na+b\,\Sigma\,(x)$ and

$\Sigma\,(xy)=a\Sigma(x)+b\Sigma(x^2)$ we get,

$89.5=7a+28b$..................(*i*)

$368.8=28a+140b$...............(*ii*)

Multiplying the first equation by 4 and subtracting it from the second
$28b=10.8$ or $b=.386$

Substituting the value in equation (*i*) we get

$7a=78.7$ or $a=11.243$

The equation $y=a+bx=y=11.243+.386x$

On this basis the value of y when $x=1$ would be $11.243+(.386\times17)=11$

Similarly other values of y (when x is 2, 3, 4, etc.) have been calculate
shown in the last column.

The above question is now solved with a short-cut method.

Year	Production y	Deviation from middle year x	x^2	xy	Computed $y=12.786+$ x
1965	12.7	-3	9	-38.1	11.628
1966	10.1	-2	4	-20.2	12.014
1967	13.0	-1	1	-13.0	12.400
1968	13.2	0	0	0	12.786
1969	12.6	$+1$	1	$+12.6$	13.172
1970	14.2	$+2$	4	$+28.4$	13.558
1971	13.7	$+3$	9	$+41.1$	13.944
Total	89.5	0	28	$+10.8$	

Two normal equations are :

$\Sigma\,(y)=na+b\Sigma(x)$ and

$\Sigma(xy)=a\Sigma(x)+b\Sigma(x^2)$

Substituting the values we get,

$89.5=7a$ or $a=12.786$

$10.8=28b$ or $b=.386$

The equation of the line would be $y=12.786+.386x$

Now the value of y when $x=-3$ would be

$$12.786+(-3\times.386)=11.628$$

Similarly other values have been calculated and shown in the last column.

PROBLEM 7.11. *Study the short-time fluctuations of the following temperatures measured in degrees Fahrenheit :—*

Date	*1971*	*Temp.*	*Date*	*1971*	*Temp.*
Feb.	*1*	*40*	*Feb.*	*11*	*78*
	2	*50*		*12*	*80*
	3	*44*		*13*	*60*
	4	*70*		*14*	*64*
	5	*52*		*15*	*62*
	6	*44*		*16*	*68*
	7	*36*		*17*	*86*
	8	*40*		*18*	*96*
	9	*56*		*19*	*94*
	10	*68*		*20*	*78*

SOLUTION. Computation of the short-time fluctuations of the temperatures measured in degrees Fahrenheit from February 1 to February 20.

Date		Temperature in degrees Fahrenheit	Add in Sevens	7 days moving average (Trend)	Short-time fluctuations
1971					
February	1	40			
	2	50			
	3	44			
	4	70	336	48	+22
	5	52	336	48	+ 4
	6	44	342	49	— 5
	7	36	366	52	—16
	8	40	374	53	—13
	9	56	402	57	— 1
	10	68	418	60	+ 8
	11	78	446	64	+14
	12	80	468	67	+13
	13	60	480	69	— 9
	14	64	498	71	— 7
	15	62	516	74	—12
	16	68	530	76	— 8
	17	86	548	78	+ 8
	18	96			
	19	94			
	20	78			

PROBLEM 7.12. *Using the data given below, explain clearly how you woul determine fluctuations in a time series :—*

Year	*Summer*	*Monsoon*	*Autumn*	*Winter*
1	*30*	*81*	*62*	*119*
2	*33*	*104*	*86*	*171*
3	*42*	*153*	*99*	*221*
4	*56*	*172*	*129*	*235*
5	*67*	*201*	*136*	*302*

SOLUTION. Computation of the short-time fluctuations.

Year and the season (1)	Series (2)	Add in fours (3)	Add in pairs (4)	Divide by (8) 4 seasons Moving Average (Trend) (5)	Fluctuations (6)
1. Summer	30				
Monsoon	81	292			
Autumn	62	295	587	73	—11
Winter	119	318	613	77	+42
2. Summer	33	342	660	83	—50
Monsoon	104	394	736	92	+12
Autumn	86	403	797	100	—14
Winter	171	452	855	107	+64
3. Summer	42	465	917	115	—73
Monsoon	153	515	980	123	+30
Autumn	99	529	1044	131	—32
Winter	221	548	1077	135	+86
4. Summer	556	578	1126	141	—85
Monsoon	172	592	1170	146	+26
Autumn	129	603	1195	149	—20
Winter	235	632	1235	154	+81
5. Summer	67	639	1271	159	—92
Monsoon	201	706	1345	168	+33
Autumn	136				
Winter	302				

The fluctuations in column (6) include both regular and irregular fluctuations. To separate the former from the total fluctuations, the following analysis table is prepared :

Calculation of the regular seasonal movement.

Year	Season			
	Summer	Monsoon	Autumn	Winter
1	...	...	—11	+42
2	—50	+12	—14	+64
3	—73	+30	—32	+86
4	—85	+26	—20	+81
5	—92	+33	...	...
Total	—300	+101	—77	+273
Average	—75	+ 25	—19	+68

Thus, the seasonal fluctuations are :

Summer	—75
Monsoon	+25
Autumn	—19
Winter	+68

PROBLEM 7.13. *Analyse the following figures of output of coal in Great Britain so as to arrive at the extent of* :

(*a*) *seasonal movement*
(*b*) *irregular fluctuations* :

Great Britain		*Quarterly output of coal (Million Tons)*
Year	*Quarters*	
1967	*I*	*68.3*
	II	*62.6*
	III	*61.1*
	IV	*63.3*
1968	*I*	*65.4*
	II	*57.9*
	III	*56.4*
	IV	*61.5*
1969	*I*	*68.1*
	II	*62.7*
	III	*62.8*
	IV	*67.0*
1970	*I*	*70.1*
	II	*59.1*
	III	*56.3*
	IV	*61.6*
1971	*I*	*59.5*
	II	*54.8*
	III	*51.1*
	IV	*58.0*

SOLUTION.

Table A

Computation of the seasonal movement and the irregular fluctuations in th output of coal in Great Britain.

Year	Quarters 1	Quarterly output of coal in million tons 2	Add in fours 3	Add in pairs 4	(Divide by 8) 4 quarterly moving average Trend 5	Short-time fluctuations 6	Seasonal movement from Table B 7	Irregula fluctuations 8
1967	I	68.3						
	II	62.6						
	III	61.1	255.3	507.7	63.5	—2.4	-3.3	+ .9

	IV	63.3	252.4	500.1	62.5	+ .8	+1.4	− .6
1968	I	63.4	247.7	490.7	61.3	+4.1	+4.4	− .3
	II	57.9	243.0	484.2	60.5	−2.6	−2.3	− .3
	III	56.4	241.2	485.1	60.6	−4.2	−3.3	+ .9
	IV	61.5	243.9	492.6	61.6	− .1	+1.4	−1.3
1969	I	68.1	248.7	503.8	63.0	+5.1	+4.4	+ .7
	II	62.7	255.1	515.7	64.5	−1.8	−2.3	+ .5
	III	62.8	260.6	523.2	65.4	−2.6	−3.3	+ .7
	IV	67.0	262.6	521.6	65.2	+1.8	+1.4	+ .4
1970	I	70.1	259.0	511.5	63.9	+6.2	+4.4	+1.8
	II	59.1	252.5	499.6	62.5	−3.4	−2.3	−1.1
	III	56.3	247.1	483.6	60.4	−4.1	−3.3	− .8
	IV	61.6	236.5	468.7	58.6	+3.0	+1.4	−1.6
1971	I	59.5	232.2	459.2	57.4	+2.1	+4.4	−2.3
	II	54.8	227.0	450.4	56.3	−1.5	−2.3	+ .8
	III	51.1	223.4					
	IV	58.0						

Table B

Calculation of the seasonal fluctuations from the short-time fluctuations of column (7)

Year	Quarters			
	I	II	III	IV
1967	...	...	−2.4	+ .8
1968	+4.1	−2.6	−4.2	− .1
1969	+5.1	−1.8	−2.6	+1.8
1970	+6.2	−3.4	−4.1	+3.0
1971	+2.1	−1.5	...	...
Total	+17.5	−9.3	−13.3	+5.5
Average	+ 4.4	−2.3	−3.3	+1.4

PROBLEM 7.14. *The following table gives monthly cost of living index numbers in the U.K. from 1967 to 1972. Calculate the season variation for each month during this period.*

	Jan.	*Feb.*	*March*	*April*	*May*	*June*	*July*	*Aug.*	*Sept.*	*Oct.*	*Nov.*	*Dec.*
1967	*175*	*172*	*171*	*165*	*164*	*163*	*166*	*164*	*165*	*167*	*169*	*169*
1968	*168*	*166*	*164*	*164*	*164*	*165*	*165*	*165*	*165*	*166*	*167*	*168*
1969	*167*	*165*	*166*	*162*	*161*	*160*	*161*	*163*	*164*	*165*	*167*	*167*
1970	*166*	*164*	*161*	*157*	*155*	*154*	*155*	*157*	*157*	*156*	*157*	*155*
1971	*153*	*152*	*150*	*147*	*147*	*145*	*147*	*145*	*145*	*145*	*146*	*148*
1972	*147*	*147*	*146*	*144*	*143*	*142*	*143*	*141*	*141*	*143*	*143*	*143*

SOLUTION. Computation of the short-time fluctuations in the monthly cost of living index numbers in the U. K. from 1967 to 1972.

Year and the months	Index	12-monthly moving average (Trend)	Short time fluctuations (col. 2 col. 3)	Year and the months	Index	12-monthly moving average (Trend)	Short time fluctuations (col. 2 col. 3)
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1967				1970			
Jan.	175			Jan.	166	161.7	+4.3
Feb.	172			Feb.	164	161.2	+2.8
March	171			March	161	160.7	+0.3
April	165			April	157	160.0	—3.0
May	164			May	155	159.3	—4.3
June	163			June	154	158.3	—4.3
July	166	167.2	—1.2	July	155	157.3	—2.3
Aug.	164	166.7	—2.7	Aug.	157	156.2	+0.8
Sept.	165	166.1	—1.1	Sept.	157	155.3	+1.7
Oct.	167	165.8	+1.2	Oct.	156	154.4	+1.6
Nov.	169	165.8	+3.2	Nov.	157	153.7	+3.3
Dec.	169	165.8	+3.2	Dec.	155	153.0	+2.0
1968				1971			
Jan.	168	165.9	+2.1	Jan.	153	152.2	+0.8
Feb.	166	165.9	+0.1	Feb.	152	151.4	+0.6
March	164	165.9	—1.9	March	150	150.4	—0.4
April	164	165.9	—1.9	April	147	149.5	—2.5
May	164	165.8	—1.8	May	147	148.5	—1.5
June	165	165.6	—0.6	June	145	147.8	—2.8
July	165	165.5	—0.5	July	147	147.2	—0.2
Aug.	165	165.5	—0.5	Aug.	145	146.8	—1.8
Sept.	165	165.5	—0.5	Sept.	145	146.4	—1.4
Oct.	166	165.5	+0.5	Oct.	145	146.1	—1.1
Nov.	167	165.3	+1.7	Nov.	146	145.8	+0.2
Dec.	168	165.0	+3.0	Dec.	148	145.5	+2.5
1969				1972			
Jan.	167	164.6	+2.4	Jan.	147	145.2	+1.8
Feb.	165	164.3	+0.7	Feb.	147	144.9	+2.1
March	166	164.2	+1.8	March	146	144.6	+1.4
April	162	164.1	—2.1	April	144	144.3	—0.3
May	161	164.1	—3.1	May	143	144.1	—1.1
June	160	164.0	—4.0	June	142	143.8	—1.8
July	161	164.0	—3.0	July	143		
Aug.	163	163.9	—0.9	Aug.	141		
Sept.	164	163.6	+0.4	Sept.	141		
Oct.	165	163.2	+1.8	Oct.	143		
Nov.	167	162.8	+4.2	Nov.	143		
Dec.	167	162.2	+4.8	Dec.	143		

Calculation of the seasonal variation for each month during the period 1967-72.

Months	Years						Total	Regular seasonal variation
	1967	1968	1969	1970	1971	1972		
Jan.		+2.1	+2.4	+4.3	+0.8	+1.8	+11.4	+2.3
Feb.		+0.1	+0.7	+2.8	+0.6	+2.1	+ 6.3	+1.3
Mar.		—1.9	+1.8	+0.3	—0.4	+1.4	+ 1.2	—0.2
April		—1.9	—2.1	—3.0	—2.5	—0.3	— 9.8	—2.0
May		—1.8	—3.1	—4.3	—1.5	—1.1	—11.8	—2.4
June		—0.6	—4.0	—4.3	—2.8	—1.8	—13.5	—2.7
July	—1.2	—0.5	—3.0	—2.3	—0.2		— 7.2	—1.4
Aug.	—2.7	—0.5	—0.9	+0.8	—1.8		— 5.1	—1.0
Sept.	—1.1	—0.5	+0.4	+1.7	—1.4		— 0.9	—0.2
Oct.	+1.2	+0.5	+1.8	+1.6	—1.1		+ 4.0	+0.8
Nov.	+3.2	+1.7	+4.2	+3.3	+0.2		+12.6	+2.5
Dec.	+3.2	+3.0	+4.8	+2.0	+2.5		+15.5	+3.1

PROBLEM 7.15 *The following table gives the value of the exports of merchandise from a port in India during the year 1969-70 to 1973-74. Calculate the Seasonal Variation Index from the data given.*

In crores of Rupees

Months	*1969-70*	*1970-71*	*1971-72*	*1972-73*	*1973-74*
April	*20*	*27*	*17*	*23*	*29*
May	*20*	*26*	*18*	*26*	*28*
June	*19*	*21*	*15*	*18*	*29*
July	*26*	*19*	*17*	*23*	*25*
August	*25*	*19*	*18*	*24*	*22*
September	*30*	*21*	*19*	*20*	*23*
October	*28*	*19*	*17*	*21*	*25*
November	*29*	*17*	*19*	*27*	*26*
December	*26*	*18*	*21*	*26*	*30*
January	*29*	*18*	*22*	*28*	*36*
February	*26*	*17*	*21*	*30*	*35*
March	*30*	*18*	*26*	*31*	*40*

SOLUTION. Calculation of the Seasonal Variation Index of the value of exports of merchandise from India during the year 1969-70 to 1973-74.

Months	Value of the exports of merchandise from India in 1969—70 to 1973—74					Total of five years	Five years' monthly average (Approx.)	Percentage of monthly average to the average of monthly averages or Seasonal Variations index
	1969—70	1970—71	1971—72	1972—73	1973—74			
April	20	27	17	23	29	116	23	96
May	20	26	18	26	28	118	24	100
June	19	21	15	18	29	102	20	83

July	26	19	17	23	25	110	22	92
Aug.	25	19	18	24	22	108	22	92
Sept.	30	21	19	20	23	121	23	96
Oct.	28	19	17	21	25	113	22	92
Nov.	29	17	19	27	26	110	24	100
Dec.	26	18	21	26	30	121	24	100
Jan.	29	18	22	28	36	133	27	113
Feb.	26	17	21	30	35	129	26	108
March	30	18	26	31	40	145	29	121
Total of monthly averages ...							286	100 approx.
Average of monthly averages ... (Total divided by 12)							24	100

PROBLEM 7.16. *The following are the quarterly index numbers of industrial Production with 1970=100 (All items) published by the Board of Trade U.K. By a moving average of four calculate a quarterly index corrected for seasonal effects.*

Year	*Index*	*Year*	*Index*
1968		*1970*	
I	*106.0*	*I*	*107.6*
II	*100.4*	*II*	*100.0*
III	*97.1*	*III*	*96.5*
IV	*105.1*	*IV*	*96.0*
1969		*1971*	
I	*107.2*	*I*	*91.5*
II	*108.6*	*II*	*89.1*
III	*107.3*	*III*	*86.4*
IV	*110.5*	*IV*	*94.1*
		1972	
		I	*91.7*
		II	*91.0*
		III	*84.4*
		IV	*91.7*

SOLUTION. Computation of the quarterly index number of industrial production in the U.K. corrected for seasonal effects.

Year and the quarters (1)	Index of production (1970=100) (2)	Sum in Fours (3)	Sum in Pairs (4)	(Divide by 8) Trend (5)	Fluctuations (6)	Normal seasonal fluctuations from Table B. (7)	Index of production adjusted for seasonal changes (col. 2 and col. 7) (8)
1968							
I	106.0					+1.1	104.9
II	100.4					— .4	100.8
III	97.1	408.6	818.4	102.3	—5.2	—3.0	100.1
IV	105.1	409.8	827.8	103.5	+1.6	+2.3	102.8

1969							
I	107.2	418.0	846.2	105.8	+1.4	+1.1	106.1
II	108.6	428.2	861.8	107.7	+0.9	— .4	109.0
III	107.3	433.6	867.6	108.4	—1.1	—3.0	110.3
IV	110.5	434.0	859.4	107.4	+3.1	+2.3	108.2
1970							
I	107.6	425.4	840.0	105.0	+2.6	+1.1	106.5
II	100.0	414.6	814.7	101.8	—1.8	— .4	100.4
III	96.5	400.1	784.1	98.0	—1.5	—3.0	99.5
IV	96.0	384.0	757.1	94.6	+1.4	+2.3	93.7
1971							
I	91.5	373.1	736.1	92.0	—0.5	+1.1	90.4
II	89.1	363.0	724.1	90.5	—1.4	— .4	89.5
III	86.4	361.1	722.4	90.3	—3.9	—3.0	89.4
IV	94.1	361.3	724.5	90.6	+3.5	+2.3	91.8
1972							
I	91.7	363.2	724.4	90.6	+1.1	+1.1	90.6
II	91.0	361.2	720.0	90.0	+1.0	— .4	91.4
III	84.4	358.8				—3.0	87.4
IV	91.7					+2.3	89.4

Table B.

Calculation of the normal seasonal fluctuations in the quarterly Index numbers of industrial production in the U.K.

Year	Quarters			
	I	II	III	IV
1968			— 5.2	+1.6
1969	+1.4	+0.9	— 1.1	+3.1
1970	+2.6	—1.8	— 1.5	+1.4
1971	—0.5	—1.4	— 3.9	+3.5
1972	+1.1	+1.0		
Total	+4.6	—1.3	—11.7	+9.6
Average	+1.2	— .3	— 2.9	+2.4
Adjusted Normal Seasonal fluctuations	+1.1	— .4	— 3.0	+2.3

PROBLEM 7.17. *The following is the rainfall in centimetres in Libya for certain months.*

	Average 1941—1965	*1971*	*1972*	*1973*	*1974*
Jan.	*2.99*	*6.2*	*3.3*	*3.4*	*3.4*
Feb.	*2.57*	*1.8*	*1.7*	*3.1*	*3.4*
March	*2.67*	*0.9*	*0.5*	*1.3*	*1.5*
April	*2.12*	*1.4*	*2.2*	*1.7*	*1.8*
May	*2.30*	*3.2*	*1.5*	*3.2*	*3.0*

June	*2.44*	*2.3*	*2.5*	*3.2*	*3.4*
July	*2.87*	*2.2*	*2.8*	*2.6*	*3.1*
August	*3.35*	*3.2*	*3.2*	*2.8*	*5.4*
Sept.	*2.54*	*3.4*	*4.2*	*2.5*	*4.9*
Oct.	*3.97*	*3.4*	*4.5*	*4.1*	*1.5*
Nov.	*3.49*	*2.7*	*6.1*	*0.8*	*6.2*
Dec.	*3.92*	*2.1*	*2.8*	*4.1*	*4.0*

Using the average for the period 1941-1965 as a norm, derive monthly index number for the period 1971-74 of the rainfall "corrected for seasonability."

SOLUTION. Calculation of the Index Number for the period 1971-74 of the rainfall in Libya, corrected for seasonability with 1941-1965=100.

Months	Average rainfall in 1941-1965 (in Cms.)	1971		1972		1973		1974	
		Rainfall in Cms.	Index Numbers	Rainfall in Cms.	Index Numbers	Rainfall in Cms.	Index Numbers	Rainfall in Cms.	Index Numbers
Jan.	2.99	6.2	207	3.3	110	3.4	114	3.4	114
Feb.	2.57	1.8	70	1.7	66	3.1	121	3.4	132
March	2.67	0.9	34	0.5	19	1.3	49	1.5	56
April	2.12	1.4	66	2.2	104	1.7	80	1.8	85
May	2.30	3.2	139	1.5	65	3.2	139	3.0	130
June	2.44	2.3	94	2.5	102	3.2	131	3.4	139
July	2.87	2.2	77	2.8	98	2.6	91	3.1	108
Aug.	3.35	3.2	96	3.2	96	2.8	84	5.4	161
Sept.	2.54	3.4	134	4.2	165	2.5	98	4.9	193
Oct.	3.97	3.4	86	4.5	113	4.1	103	1.5	38
Nov.	3.49	2.7	77	6.1	175	0.8	23	6.2	178
Dec.	3.92	2.1	54	2.8	71	4.1	105	4.0	102

PROBLEM 7.18. *The following is an index no. of the price of lead from 1956-1975 together with the "Statist" wholesale price index of the period. Construct an index of lead prices, "corrected" for changes in the wholesale price level.*

Year	*Index of wholesale price*	*Index No. of Lead Prices*
1956	*125*	*157*
1957	*122*	*125*
1958	*119*	*109*
1959	*114*	*117*
1960	*96*	*95*
1961	*82*	*71*
1962	*79*	*63*
1963	*78*	*65*
1964	*81*	*61*
1965	*83*	*78*

1966	*88*	*95*
1967	*102*	*121*
1968	*90*	*83*
1969	*94*	*85*
1970	*128*	*127*
1971	*142*	*129*
1972	*151*	*129*
1973	*155*	*129*
1974	*160*	*129*
1975	*164*	*142*

SOLUTION. Computation of the Index numbers of the price of lead from 1956-1975, corrected for changes in the wholesale price level.

Year	Index Number of wholesale price	Index Number of price of lead	Corrected index number of the price of lead	
1956	125	157	157 × 100/125	126
1957	122	125	125 × 100/122	102
1958	119	109	109 × 100/119	92
1959	114	117	117 × 100/114	103
1960	96	95	95 × 100/ 96	99
1961	82	71	and so on	87
1962	79	63		80
1963	78	65		83
1964	81	61		75
1965	83	78		94
1966	88	95		108
1967	102	121		119
1968	90	83		92
1969	94	85		90
1970	128	127		99
1971	142	129		91
1972	151	129		85
1973	155	129		83
1974	160	129		81
1975	164	142		87

PROBLEM 7.19. *The number (in hundreds) of letters posted in a certain city on each day in a typical period of five weeks was as follows :—*

	Sun.	*Mon.*	*Tues.*	*Wed.*	*Thurs.*	*Fri.*	*Sat.*	*Total for each week*
1st week	*18*	*161*	*170*	*164*	*153*	*181*	*76*	*923*
2nd week	*18*	*165*	*179*	*157*	*168*	*195*	*85*	*967*
3rd week	*21*	*162*	*169*	*153*	*139*	*185*	*82*	*911*
4th week	*24*	*171*	*182*	*170*	*162*	*179*	*95*	*983*
5th week	*27*	*162*	*186*	*170*	*170*	*182*	*120*	*1017*
Total for all weeks	*108*	*821*	*886*	*814*	*792*	*922*	*458*	*4801*

Calculate the 'average fluctuations' indices within a week.

SOLUTION. Calculation of the indices of average fluctuations within a week.

Days	1st week	2nd week	3rd week	4th week	5th week	Total	Weekly Average	Indices of weekly average (137=100)	Percentage Fluctuations
Sunday	18	18	21	24	27	108	22	16	−84
Monday	161	165	162	171	162	821	164	120	+20
Tuesday	170	179	169	182	186	886	177	129	+29
Wed.	164	157	153	170	170	814	163	119	+19
Thursday	153	168	139	162	170	792	158	115	+15
Friday	181	195	185	179	182	922	184	134	+34
Saturday	76	85	82	95	120	458	92	67	−33
						Total weekly average	960		
						Average of weekly average	137		

Note :—Fluctuations can be eliminated from the data either with the method of moving average or the method of least squares.

REVISIONAL PROBLEMS–7

7.20. Apply the method of semi-averages for determining trend of the following data and estimate the value for 1970 :

Years	Sales (in thousand units)	Years	Sales (in thousand units)
1963	20	1966	30
1964	24	1967	28
1965	22	1968	32

If the actual figure of sales for 1970 is 35.000 units, how do you account for the difference between the figure you obtain and the actual figure given to you.

7.21. Find the centred 4 year moving averages for the following time series data :

Years	1960	1961	1962	1963	1964	1965	1966	1967
y	30.1	45.4	39.3	41.4	42.2	46.4	46.6	49.2

[*Ans*	*1962*	*1963*	*1964*	*1965*
	40.56	*42.20*	*43.24*	*45,12*]

7.22. Compute the trend value by the method of least squares from the data given below :

Year	1962	1963	1964	1965	1966	1967	1968	1969
No. of sheep (in lakhs)	56	55	51	47	42	38	35	32

(y=*44.5*--*1.86*x)

7.23: Fit a straight line trend by the method of least squares.

Years	Milk Consumption (million gallons)	Years	Milk Consumption (million gallons)
1960	102.3	1964	114.8
1961	101.9	1965	118.7
1962	105.8	1966	124.5
1963	112.0	1967	129.9
		1968	134.8

[y=*116.1*+*3.4*x]

7.24. The prices of a commodity during 1960—65 are given below. Fit a parabola $y=a+bx+cx^2$ to these data. Estimate the price of the commodity for the year 1966 :

Year	Prices	Year	Prices
1960	100	1963	140
1961	107	1964	181
1962	128	1965	192

Also plot the actual and trend values on graph.

[*Ans*: *Trend* : *88.82, 105.07, 124.88, 138.20, 175.30 and 206.70*]

7.25. The following table gives quarterly expenditure data over a number of years. Obtain seasonal corrections for the data :

Quarter	1970	1971	1972	1973
I	78	84	92	100
II	62	64	70	81
III	56	61	63	72
IV	71	82	83	96

7.26. The trend of the annual sales of the Bharat Aluminium Company is described by the following equation :

$$Yc=12+0.7x$$

where 1970=0

x unit=1 year

and y unit=annual production.

Step the equation down to a month to month basis, and shift the origin to January, 1970.

7.27. A store has an annual trend equation for its sales index at

$$Y_{1957}=100+3x$$

If its December seasonal is 200 and its July seasonal 60, which was a relatively better month for sales, July 1960 when the sales index was at 70 or December 1958 when the index was at 220 ?

7.28. Obtain the seasonal fluctuations from the following time series : Quarterly output of coal for years :

Years	Quarters			
	I	II	III	IV
1965	65	58	56	61
1966	68	63	63	67
1967	70	59	56	52
1968	60	55	51	58

7.29. Find trend values (mixed with cyclical movements, if any) from the following data of output, by the method of moving averages :

Quarters	Years			
	1965	1966	1967	1968
I	29	40	47	45
II	37	42	51	49
III	43	55	63	60
IV	34	43	53	48

Interpolation

8

Interpolation means the insertion of the most likely figure, in a continuous series. Interpolation is based on two assumptions. Firstly, that the quantity changes continuously without any break, and secondly, that the rate of change is uniform, and there are no sudden jumps in the data. If these assumptions are not fulfilled the interpolated figure would be a fictitious and unlikely one. In some problems that have been solved here such a situation has arisen, and we have given brief notes at the end of such questions. Questions have been solved mostly by important methods of interpolation, *viz.*, Parabolic Curve Method, Newton's Method, Newton-Gauss Method, Sterling Formula, Binomial Expansion Method, Lagrange's Method and Method of Divided Differences. Few questions have been solved by other methods also.

PARABOLIC CURVE METHOD

PROBLEM 8.1. *The following table gives the death rate per 100,000 population on account of T. B. in a certain district of Uttar Pradesh.*

Year	*Death rate*
1966	*160*
1968	*175*
1970	*180*

Estimate the death rate for 1969.

SOLUTION. Estimation of the death rate for 1969 by a parabolic curve.

Years		1966	1968	1969	1970
Deviation from the year 1969	x	-3	-1	0	$+1$
Death Rate	y	160	175	y_0	180

Since the known values are three, we will take a parabola of the 2nd order *viz.*

$$y=a+bx+cx^2$$

Substituting the different values of x and y, we get the following equations,

$160=a-3b+9c$ .. .. .. (*i*)

$175=a-b+c$ .. .. .. (*ii*)

$y_0=a$ .. .. .. (*iii*)

$180=a+b+c$ .. .. .. (*iv*)

To deduce the value of y_0 which will give the estimated death-rate for 1969, we proceed as follows :

Add (ii) and (iv) ;

$355=2a+2c$ .. .. .. (*v*)

Multiplying (iv) by 3,

$540=3a+3b+3c$ .. .. .. (*vi*)

Add (i) and (vi)

$700=4a+12c$ .. .. .. (*vii*)

Multiplying (v) by 6

$2130=12a+12c$ .. .. .. (*viii*)

Subtracting (vii) from (viii)

$8a=1430$

or $a=178.75$

Therefore $y_0=178.75$

Thus, the estimated death-rate for 1969 is 178.75.

PROBLEM 8.2. *The pressure of wind in pounds per square foot, corresponding to the velocity in miles per hour has been determined by experiment to be approximately as follows :—*

Velocity	*20,*	*30,*	*40,*	*50*
Pressure	*2,*	*4.4,*	*7.9,*	*13.2*

Estimate the pressure for a velocity of 35 miles per hour.

SOLUTION. Estimation of the pressure for a velocity of 35 miles per hour by the method of fitting a parabolic curve.

Velocity		20	30	35	40	50
Deviation from 35		−15	− 5	0	+ 5	+15
5 taken as common from deviations	x	− 3	− 1	0	+ 1	+ 3
Pressure	y	2	4.4	y_0	7.9	13.2

Since the known values are four, we will take a parabola of the 3rd order, *viz.*

$$y=a+bx+cx^2+dx^3$$

Substituting the given values of x and y in the above we get the following equations.

$2=a-3b+9c-27d$ .. .. .. (*i*)

$4.4=a-b+c-d$ .. .. .. (*ii*)

$y_0=a$ .. .. .. (*iii*)

$7.9=a+b+c+d$ .. .. .. (*iv*)

$13.2=a+3b+9c+27d$ .. .. .. (*v*)

To deduce the value of y_0 which will give the pressure for a velocity of 35 miles per hour, we proceed as follows :

Adding (*i*) and (*v*),

$15.2=2a+18c$ .. .. .. (*vi*)

Adding (*ii*) and (*iv*),

$12.3=2a+2c$

or $110.7=18a+18c$ .. .. .. (*vii*)

Subtracting (*vi*) and from (*vii*)

$16a=95.5$

Therefore, $y_0=5.97$

Thus, the estimated pressure for a velocity of 35 miles per hour is 5.97.

PROBLEM 8.3. *The following table shows the value of an immediate life annuity for every £ 100 paid :—*

Age in years	*Annuity (£)*
40	*6.2*
50	*7.2*
60	*9.1*
70	*12.0*

Interpolate for the age 42

SOLUTION. Interpolation of the value of a life annuity for the 42 by a parabolic curve.

Age in years		40	42	50	60	70
Deviation from (42)		−2	0	+ 8	+18	+28
Taken as common from the deviations 2	x	−1	0	+ 4	+ 9	+14
Annuity (£)	y	6.2	y_0	7.2	9.1	12.0

Since the known values are four, we will take a parabola of the 3rd order, *viz.*,

$$y=a+bx+cx^2+dx^3$$

Substituting the different values of x and y we get the following equations:—

$$6.2=a-b+c-d \quad \ldots \quad (i)$$
$$yo=a \quad \ldots \quad (ii)$$
$$7.2=a+4b+16c+64d \quad \ldots \quad (iii)$$
$$9.1=a+9b+81c+729d \quad \ldots \quad (iv)$$
$$12.0=a+14b+196c+2744d \quad \ldots \quad (v)$$

To deduce the value of y_0 which will give the estimated value of a life annuity for the age 42, we proceed as follows:—

Multiplying (*i*) by 4,

$$24.8=4a-4b+4c-4d \quad \ldots \quad (vi)$$

Adding (*iii*) and (*vi*)

$$32.0=5a+20c+60d$$
$$\text{or } 384.0=60a+240c+720d \quad \ldots \quad (vii)$$

Multiplying (*i*) by 9.

$$55.8=9a-9b+9c-9d \quad \ldots \quad (viii)$$

Adding (*iv*) and (*viii*)

$$64.9=10a+90c+720d \quad \ldots \quad (ix)$$

Subtracting (*ix*) from (*viii*)

$$319.1=50a+150c$$
$$\text{or } 20103.3=3150a+9450c \quad \ldots \quad (x)$$

Multiplying (*i*) by 14

$$86.8=14a-14b+14c-14d \quad \ldots \quad (xi)$$

Adding (*v*) and (*xi*)

$$98.8=15a+210c+2730d$$
$$\text{or } 7113.6=1080a+15120c-196560d \quad \ldots \quad (xii)$$

Multiplying (*ix*) by 273

$$17{,}717.7=2730a+24570c+196560d \quad \ldots \quad (xiii)$$

Subtracting (*xii*) from (*xiii*)

$$10{,}604.1=1650a+9450c \quad \ldots \quad (xiv)$$

Subtracting (*xiv*) from (*x*)

$$9{,}499.2=1500a$$

Therefore, $a=6{,}333$

or, $y_0=6.333.$

Thus, the estimated value of the life annuity for the age 42 is £6.333.

PROBLEM. 8.4. *The following table gives the census of population of an Indian State in 1941, 1951, 1961, and 1971.*

Estimate the population of the State in 1964, making your method clear.

Year	*Population (in thousands)*
1941	*2,797*
1951	*2,935*
1961	*3,047*
1971	*3,354*

SOLUTION. Estimation of the said state in 1924 by fitting a parabolic curve.

Year	1941	1951	1961	1964	1971
Deviation from 1924	—23	—13	— 3	0	+ 7
Population in thousands	2,797	2,935	3,047	y_0	3,354

Since the known values are four, we will take a parabola of the 3rd order, *viz.*,

$$y = a + bx + cx^2 + dx^3$$

Substituting the different values of x and y, we get the following equations:—

$2797 = a - 23b + 529c - 12167d$ .. .. .. (*i*)

$2935 = a - 13b + 169c - 2197d$ .. .. .. (*ii*)

$3047 = a - 3b + 9c - 27d$ .. .. .. (*iii*)

$y_0 = a$ .. .. .. (*iv*)

$3354 = a + 7b + 49c + 343d$ .. .. .. (*v*)

To deduce the value of y_0 which will give the estimated population of the state in 1964, we proceed as follows:

Multiplying (*i*), (*ii*) and (*v*) by 3 and the (*iii*) by 23, 13 and 7 we get

$8391 = 3a - 69b + 1587c - 36501d$ .. .. (*vi*)

$8805 = 3a - 39b + 507c - 6591d$ .. .. (*vii*)

$10062 = 3c + 21b + 147c + 1029d$ .. .. (*viii*)

$70081 = 23a - 69b + 207c - 621d$ .. .. (*ix*)

$39611 = 13a - 39b + 117c - 351d$ .. .. (*x*)

$21329 = 7a - 21b + 63c - 189d$ .. .. (*xi*)

Subtracting (*vi*) from (*ix*) and (*vii*) from (*x*),

$61690 = 20a - 1380c + 35880d$ .. .. (*xii*)

$30806 = 10a - 390c + 6240d$ .. .. (*xiii*)

Adding (*viii*) and (*xi*),

$31391 = 10a + 210c + 840d$ .. .. (*xiv*)

Multiplying (*xii*) and (*xiii*) by 7, and (*xiv*) by 46 and 13,

$431830 = 140a - 9660c + 251160d$ .. .. (*xv*)

$215642 = 70a - 2730c + 43680d$ .. .. (*xvi*)

$1443986 = 460c + 9660c + 38640d$ .. .. (*xvii*)

$408083 = 130a + 2730c + 10920d$ .. .. (*xviii*)

Adding (*xv*) and (*xvii*) and (*xvi*) and (*xviii*),

$1875816 = 600a + 289800d$ .. .. (*xix*)

$623725 = 200a + 54600d$ .. .. (*xx*)

Multiplying (*xix*) by 13 and (*xx*) by 69,

$$24385608 = 7800a + 3767400d \quad \ldots \quad (xxi)$$

$$43037025 = 13800a + 3767400d \quad \ldots \quad (xxii)$$

Subtracting (*xxi*) from (*xxii*),

$$18651417 = 6000a$$

Therefore $a = 3108.5695$

Or $= y_0 = 3108.5695$

Thus the population of the given state in 1964 is estimated at 3108.5695 (thousands).

PROBLEM 8.5. *Estimate the anuual sales of pencils for 1970 from the following records of wholesale merchants.*

Year	*Sales of pencils in lakhs of dozens*
1960	*25*
1964	*30*
1968	*40*
1972	*55*
1976	*60*

SOLUTION. Interpolation of the annual sales of pencils for 1970

Year		1960	1964	1968	1970	1972	1976
Deviation from 1970		—10	—6	—2	0	+2	+6
2 taken as common from deviation	x	— 5	—3	—1	0	+1	+3
Sales of pencils in lakhs of dozens	y	25	30	40	y_0	55	60

Since the known values are five, we will take a parabola of the 4th order, *viz.*

$$y = a + bx + cx^2 + dx^3 + ex^4$$

Substituting the given values of x and y, we get the following equations. :

$$25 = a - 5b + 25c - 125d + 625e \quad \ldots \quad (i)$$

$$30 = a - 3b + 9c - 27d + 81e \quad \ldots \quad (ii)$$

$$40 = a - b + c - d + e \quad \ldots \quad (iii)$$

$$y_0 = a \quad \ldots \quad (iv)$$

$$55 = a + b + c + d + e \quad \ldots \quad (v)$$

$$60 = a + 3b + 9c + 27d + 81c \quad \ldots \quad (vi)$$

To deduce the value of y_0, which will give the estimated annual sales of pencils for 1970, we proceed as follows :

Multiplying (v) by 5 and 3

$$275 = 5a + 5b + 5c + 5d + 5e \quad \ldots \quad (vii)$$

$$165 = 3a + 3b + 3c + 3d + 3e \quad \ldots \quad (viii)$$

Adding (i) and (vii),

$$300 = 6a + 30c - 120d + 630e \quad \ldots \quad (ix)$$

Adding (ii) and ($viii$),

$$195 = 4a + 12c - 24d + 84e$$

$$\text{or } 975 = 20a + 60c - 120d + 420e \quad \ldots \quad (x)$$

Subtracting (ix) from (x),

$$675 = 14a + 30c - 210e \quad \ldots \quad (xi)$$

Adding (*ii*) and (*vi*), and (*iii*) and (*v*),

$90 = 2a + 18c + 162e$ .. .. .. (*xii*)

$95 = 2a + 2c + 2e$ .. .. .. (*xiii*)

Multiplying (*xiii*) by 15 and 9,

$1425 = 30a + 30c + 30e$ .. .. .. (*xiv*)

$855 = 18a + 18c + 18e$ .. .. .. (*xv*)

Subtracting (*xi*) from (*xiv*) and (*xii*) from (*xv*),

$750 = 16a + 240e$ .. .. .. (*xvi*)

$765 = 16a = 144e$ .. .. .. (*xvii*)

Multiplying (*xvi*) by 3 and (*xvii*) by 5,

$2250 = 48a + 720e$ .. .. .. (*xviii*)

$3825 = 80a + 720e$ .. .. .. (*xix*)

Adding (*xviii*) and (*xix*),

$6075 = 128a$

Therefore $a = 47.46$

or $yo = 47.46$

Thus, the annual sale of pencils for 1970 is estimated at 47.46 lakhs of dozens.

NEWTON'S METHOD

PROBLEM 8.6. *State Newton's formula for interpolation for equal intervals and the assumptions underlying it. Use it to find the annual net premium at age 25 from the table given below :—*

Age	*Annual net premium*
20	*.01427*
24	*.01581*
28	*.01772*
32	*.01996*

SOLUTION. Interpolation of the annual net premium at age 25 by Newton's method.

Age (in years)		Annual Net premiums		Difference					
				First Δ^1		Second Δ^2		Third Δ^3	
20	x_0	.01427	y_0						
24	x_1	.01581	y_1	.00154	Δ^1_0				
28	x_2	.01772	y_2	.00191	Δ^1_1	.00037	Δ^2_0		
32	x_3	.01996	y_3	.00224	Δ^1_2	.00033	Δ^2_1	—.00004	Δ^3_0

$$x = \frac{\text{Year of interpolation} - \text{Year of Origin}}{\text{Time distance between adjoining years}} = \frac{25-20}{28-24} = \frac{5}{4} = 1.25$$

Substituting the above values in the Newton's formula

$$yx = y_0 + x\Delta^1_0 + \frac{x(x-1)}{1\times2}\,\Delta^2_0 + \frac{x(x-1)\,(x-2)}{1\times2\times3}\,\Delta^3_0$$

where yx represents the premium at age 25, which is to be interpolated.

We get,

$$yx = .01427 + (1.25 \times .00154) + \frac{1.25\,(1.25-1)\,(.00037)}{1 \times 2} + \frac{1.25(1.25-1)(1.25-2)(-.00004)}{1 \times 2 \times 3}$$

$$= .01427 + .001925 + .000578 + .0000016 = .0167742 = .01678 \text{ approx.}$$

Thus, the annual net premium payable at the age 25 is .01678.

PROBLEM 8.7. *The following are the annual premiums in a certain Life Insurance Company for a policy of Rs. 500 payable at the death with an agreed bonus :—*

Age next birthday	*Annual Premium*
	Rs. ps.
25	*24—62*
30	*27—69*
35	*31—56*
40	*36—37*
45	*42—31*

Calculate the premium at age 36.

SOLUTION. Annual premium on a life policy of Rs. 500, and their differences.

Age next birthday (x)	Annual premium in Paisas	Differences			
		First Δ^1	Second Δ^2	Third Δ^3	Fourth Δ^4
25 x_0	2462 y_0	307 Δ^1_0			
30 x_1	2769 y_1	387 Δ^1_1	80 Δ^2_0		
35 x_2	3156 y_2	481 Δ^1_2	94 Δ^2_1	14 Δ^3_0	
40 x_3	3637 y_3	594 Δ^1_3	113 Δ^2_2	19 Δ^3_1	5 Δ^4_0
45 x_4	4231 y_4				

$$x = \frac{36-25}{30-25} = \frac{11}{5} = 2.2$$

Substituting the above values in the Newton's formula for interpolation.

$$yx = y_0 + x\Delta^1{}_0 + \frac{x(x-1)}{1 \times 2}\Delta^2{}_0 + \frac{x(x-1)(x-2)}{1 \times 2 \times 3}\Delta^3{}_0 + \frac{x(x-1)(x-2)(x-3)}{1 \times 2 \times 3 \times 4}\Delta^4{}_0$$

Where yx represents the premium at age 36, which is to be interpolated.

We get,

$$yx = 2462 + (2.2 \times 307) + \frac{2.2\,(2.2-1)\,(80)}{1 \times 2} + \frac{2.2\,(2.2-1)\,(2.2-2)\,(14)}{1 \times 2 \times 3} + \frac{2.2\,(2.2-1)\,(2.2-2)\,(2.2-3)\,(5)}{1 \times 2 \times 3 \times 4}$$

Or $yx = 2462 + 675.4 + 105.6 + 1.232 - .088$

Or $yx = 3244.144$ paisa.

Thus, the premium at age 36 for the said policy is estimated at Rs. 32.44 paisa

PROBLEM 8.8 *The following table gives the quantities of a certain of tea demanded at prices noted against each. Estimate the probable demand the price is Rs. 1.87 per packet.*

Price of tea per packet	*Quantity demanded in thousands of Packets*
Rs.	
1.25	*82.5*
1.50	*70.8*
1.75	*63.1*
2.00	*55.0*
2.25	*48.9*

SOLUTION. Quantity of tea demanded at given prices and differences.

Price of tea per packet (x)		Quantity demanded in thousands of packets		Differences			
				First Δ^1	Second Δ^1	Third Δ^2	Fou Δ
Rs.							
1.25	x_0	82.5	y_0	$-11.7\ \Delta^1_0$			
1,50	x_1	70.8	y_1	$-7.7\ \Delta^1_1$	$+4.0\ \Delta^2_0$		
1.75	x_2	63.1	y_2	$-8.1\ \Delta^1_2$	$-4\ \Delta^2_1$	$-4.4\ \Delta^3_0$	
2.00	x_3	55.0	y_3	$-6.1\ \Delta^1_3$	$+2.0\ \Delta^2_2$	$+2.4\ \Delta^3_1$	$+6.8\ \Delta$
2.25	x_4	48.9	y_4				

$$x=\frac{(\text{Rs. }1.87)-(\text{Rs.}1.25)}{(\text{Rs. }2.00)-(\text{Rs.}1.75)}=\frac{62}{25}=2.48 \textbf{ paisa}$$

Substituting the above value in the Newton's formula of interpolation,

$$yx=y_0+x\triangle^1_0+\frac{x(x-1)}{1\times2}\triangle^8_0+\frac{x(x-1)(x-2)}{1\times2\times3}\triangle^3_0+\frac{x(x-1)(x-2)(x-3)}{1\times2\times3\times4}\triangle^4_0$$

Where yx represents the thing to be interpolated, that is the demand at price of Rs. 1.87 in this case.

We get,

$$yx=82.5+(2.48\times-11.7)+\frac{2.48(2.5-1)\times4}{1\times2}+\frac{2.48(2.5-1)(2.5-2)\times-4.4}{1\times2\times3}+\frac{2.48(2.5-1)(2.5-2)(2.5-3)\times6.8}{1\times2\times3\times4}$$

or $yx=82.5-29.01+7.44-1.364-.2633$

or $yx=59.31$ thousand packets

Thus, the estimated demand of tea at a price of Rs. 1.87 is 59.31 thousand packets.

PROBLEM 8.9. *The gross profits of the Bulsar Sugar Co., Ltd. are given below :—*

Years	*Gross profit (in lakhs of rupees)*
1965—66	*4.86*
1967—68	*12.64*
1969—70	*13.68*
1971—72	*16.65*
1973—74	*23.29*

Make an estimate for 1972-73 and 1974-75.

SOLUTION. Gross profit of the Bulsar Sugar Co., Ltd., for the given ears and their differences.

Years (x)		Gross profit (in lakhs of rupees) (y)		Differences			
				First Δ^1	Second Δ^2	Third Δ^3	Fourth Δ^4
965-66	x_0	4.86	y_0	7.78 Δ^1_0			
967-68	x_1	12.64	y_1	1.04 Δ^1_1	—6.74 Δ^2_0		
969-70	x_2	13.68	y_2	2.97 Δ^1_2	1.93 Δ^2_1	8.67 Δ^3_0	—6.93 Δ^4_0
971-72	x_3	16. 65	y_3	6.64 Δ^1_3	3.67 Δ^2_2	1.74 Δ^3_1	
973-74	x_4	23.29	y_4				

$$x \text{ for } 1972\text{-}73 = \frac{\text{Year of interpolation — Year of Origin}}{\text{Time distance between adjoining years}}$$

$$= \frac{(1972\text{-}73)—(1965\text{-}66)}{(1973\text{-}74)—(1971\text{-}72)} = \frac{7}{2} = 3.5$$

$$x = \text{for } 1974\text{-}75 = \frac{\text{Year of extrapolation -- Year of Origin}}{\text{Time distance between adjoining years}}$$

$$= \frac{(1974\text{-}75)—(1965\text{-}66)}{(1975\text{-}76)—(1973\text{-}74)} = \frac{9}{2} = 4.5$$

Substituting the given values in the Newton's formula

$$yx = y_0 + x\Delta^1_0 + \frac{x(x-1)}{1\times 2}\Delta^2_0 + \frac{x(x-1)(x-2)}{1\times 2\times 3}\Delta^3_0 + \frac{x(x-1)(x-2)(x-3)}{1\times 2\times 3\times 4}\Delta^4_0$$

Where yx represents the thing to be interpolated or extrapolated e get,

(*i*) *Estimate of the profits for the year 1972-73.*

$$yx\ (1972\text{-}73) = 4.86 + (3.5\times 7.78) + \frac{3.5\ (3.5--1)\times -6.74)}{1\times 2}$$

$$+ \frac{3.5\ (3.5—1)\ (3.5—2)\times 8.67,}{1\times 2\times 3} + \frac{3.5\ (3.5—1)\ (3.5—2)\ (3.5—3)\times -6.93}{1\times 2\times 3\times 4}$$

or $yx\ (1972\text{-}73) = 4.86 + 27.23 -- 29.488 + 18.9656 — 1.894$

or $yx\ (1972\text{-}73) = 19.6736$ lakhs of rupees.

(*ii*) *Estimate of the profit for the year 1944-45.*

$$yx\ (1974\text{-}75) = 4.86 + (4.5\times 7.78) + \frac{4.5\ (4.5—1)\times —6.74}{1\times 2}$$

$$+ \frac{4.5\ (4.5—1)\ (4.5—2)\times 8.67}{1\times 2\times 3} + \frac{4.5\ (4.5—1)\ (4.5—2)\ (4.5—3)\times —6.93}{1\times 2\times 3\times 4}$$

or $yx\ (1974\text{-}75) = 4.86 + 35.01 -\cdot 53.0775 + 56.897 — 17.0543.$

or $yx\ (1974\text{-}75) = 26.635$ lakhs of rupees.

Thus the estimated gross profits of the Bulsar Sugar Co. Ltd. for 1972-73 nd 1974-75 are Rs. 19.6736 or (19.70) lakhs and Rs. 26.635 or (26-65) lakhs spectively.

PROBLEM 8.10. *Estimate by Newton's method of interpolation, the expectaon of life at age 22 from the following data, stating the assumption underlying the rmula used by you :—*

Age	*10*	*15*	*20*	*25*	*30*	*35*
pectation of life (in years)	*35.4*	*32.2*	*29.1*	*26.0*	*23,1*	*20.4*

SOLUTION. Interpolation of the expectation of life at age 22 by Newt method.

Age (in years) (x)	Expectation of life (in years) (y)	Differences				
		Δ^1	Δ^2	Δ^3	Δ^4	Δ
10 x_0	35.4 y_0	-3.2 Δ^1_0	$.1$ Δ^2_0	$-.1$ Δ^3_0	$.3$ Δ^4_0	$-.5$
15 x^1	32.2 y_1	-3.1 Δ^1_1	$.0$ Δ^2_1	$.2$ Δ^3_1	$-.2$ Δ^4_1	
20 x_2	29.1 y_2	-3.1 Δ^1_2	$.2$ Δ^2_2	$.0$ Δ^3_2		
25 x_3	26.0 y_3	-2.9 Δ^1_3	$.2$ Δ^2_3			
30 x_4	23.1 y_4	-2.7 Δ^1_4				
35 x_5	20.4 y_5					

$$x = \frac{\text{Year of interpolation—Year of Origin}}{\text{Time Distance between adjoining years}} = \frac{22-10}{25-20} = \frac{12}{5} = 2.4$$

Substituting the given values in the Newton's formula of interpolation.

$$yx = y_0 + x\Delta^1_0 + \frac{x(x-1)}{1\times2}\Delta^2_0 + \frac{x(x-1)(x-2)}{1\times2\times3}\Delta^3_0$$

$$+\frac{x(x-1)(x-2)(x-3)}{1\times2\times3\times4}\Delta^4_0 + \frac{x(x-1)(x-2)(x-3)(x-4)}{1\times2\times3\times4\times5}\Delta^5_0$$

Where yx represents the thing to be interpolated, that is, the expectation life at age 22 in this case.

We get,

$$yx = 35.4 + (2.4\times-3.2) + \frac{2.4(2.4-1)\times.1}{1\times2} + \frac{2.4(2.4-1)(2.4-2)\times-.1}{1\times2\times3}$$

$$+\frac{2.4(2.4-1)(2.4-2)(2.4-3)\times.3}{1\times2\times3\times4}$$

$$+\frac{2.4(2.4-1)(2.4-2)(2.4-3)(2.4-4)\times-.5}{1\times2\times3\times4\times5}$$

or $yx = 35.4 - 7.68 + .168 - .0226 - .01008 - .005376$

Or $yx = 27.85$ years.

Thus, the expectation of life at age 22 is 27.85 years.

PROBLEM 8.11. *If lx represents the numbers living at age x in a life tab find as accurately as the data will permit, lx for values of $x = 35, 42,$ and given.*

$l_{20} = 512$: $l_{30} = 439$; $l_{40} = 346$; $l_{50} = 243$.

SOLUTION. Interpolation of numbers living at ages 35, 42 and Newton's Method.

Age (in years) (x)	Numbers living lx (y)	Differences		
		First Δ^1	Second Δ^2	Third Δ^3
20 x_0	512 y_0	-73 Δ^1_0	-20 Δ^2_0	$+10$ Δ^3
30 x_1	439 y_1	-93 Δ^1_1	-10 Δ^2_1	
40 x_2	346 y_2	-103 Δ^1_2		
50 x_3	243 y_3			

$$x=\frac{\text{Year of Interpolation—Year of Origin}}{\text{Time Distance between adjoining years}}$$

Substituting the given values, we get

$$x_{35}=\frac{35-20}{40-30}=\frac{15}{10}=1.5$$

$$x_{42}=\frac{42-20}{50-40}=\frac{22}{10}=2.2$$

$$x_{47}=\frac{47-20}{50-40}=\frac{27}{10}=2.7$$

Applying to the above data, the Newton's formula

$$yx=y_0+\triangle^1{}_0+\frac{x(x-1)}{1\times2}\triangle_0^2+\frac{x(-1)(x-2)}{1\times2\times3}\triangle^3{}_0$$

Where yx represents the thing to be interpolated, that is lx in this case,

We get,

(1) yx_{35} or $l_{35}=512+(1.5\times-73)+\frac{1.5\,(1.5-1)\times-20}{1\times2}$

$$+\frac{1.5\,(1.5-1)\,(1.5-2)\times10}{1\times2\times3}$$

or $l_{35}=512-109{\cdot}50-7.50-.625$ or $l_{35}=394.375$

(2) yx_{42} or $l_{42}=512+(2.2\times-73)+\frac{2.2\,(2.2-1)-20}{1\times2}$

$$+\frac{2.2\,(2.2-1)\,(2.2-2)\times10}{1\times2\times3}$$

or $l_{42}=512-160.6-26.4+.88$ or $l_{42}=325.88$ or 326.

(3) yx_{47} or $l_{47}=512+(2.7\times-73)+\frac{2.7\,(2.7-1)\times-20}{1\times2}$

$$+\frac{2.7\,(2.7-1)(2.7-2)\times10}{1\times2\times3}$$

or $l_{47}=512-197.1-45.9+5.355$ or $l_{47}=274.355$

Thus, the interpolated values of l_{35}, l_{42} and l_{47} are 394, 326 and 274 respectively.

PROBLEM 8.12. *Estimate the expectation of life at the age of 16 years using the following data.*

Age in years	*Expectation of life*
10	*35.4*
15	*32.3*
20	*29.2*
25	*26.0*
30	*32.2*
35	*20.4*

SOLUTION. Interpolation of the expectation of life at age 16 by New method.

Age in years	Expectation of life	Differences First Δ^1	Second Δ^2	Third Δ^3	Fourth Δ^4	Fi
10 x_0	35.4 y_0					
		$-3.1\Delta^1_0$				
15 x_1	32.3 y_1		$+ .0\Delta^2_0$			
		$-3.1\Delta^1_1$		$- .1\Delta^3_0$		
20 x_2	29.2 y_2		$- .1\Delta^2_1$		$+ 9.6\Delta^4_0$	
		$-3.2\Delta^1_2$		$+ 9.5\Delta^3_1$		—46.5
25 x_3	26.0 y_3		$+ 9.4\Delta^2_2$		$-36.9\Delta^4_1$	
		$+6.2\Delta^1_3$		$-27.4\Delta^3_2$		
30 x_4	32.2 y_4		$-18.0\Delta^2_3$			
		$-11.8\Delta^1_4$				
35 x_5	20.4 y_5					

$$x = \frac{\text{Year of interpolation—year of origin}}{\text{Time distance between adjoining years}} = \frac{16-10}{20-15} - \frac{6}{5} = 1.2$$

Substituting the given values in the Newton's formula of interpolation.

$$yx = yo + x\Delta^1_0 + \frac{x(x-1)}{1\times2}\Delta^2_0 + \frac{x(-1)(x-2)}{1\times2\times3}\Delta^3_0$$
$$+\frac{x(x-1)(x-2)(x-3)}{1\times2\times3\times4}\Delta^4_0$$
$$\frac{x(x-1)(x-2)(x-3)(x-4)}{1\times2\times3\times4\times5}\Delta^5_0$$

Where yx represents the thing to be interpolated, that is, the expectatio life at age 16 in this case.

We get,

$$yx = 35.4 + (1.2\times-3.1) + \frac{1.2(1{\cdot}2-1)\times.0}{1\times2} + \frac{1.2(1.2-1)(1.2-2)\times-.1}{1\times2\times3}$$
$$+\frac{1.2(1.2-1)(1.2-2)(1.2-3)\times9.6}{1\times2\times3\times4}$$
$$+\frac{1.2(1.2-1)(1.2-2)(1.2-3)(1.2-4)\times-46.5}{1\times2\times3\times4\times5}$$

or $yx = 35.4 - 3.72 - 0 + .0032 + .13824 + .374976$

or $yx = 32.196416$

Thus, the expectation of life at age 16 is 32.20 years.

PROBLEM 8.13. *From the following data, estimate the number of pers earning weekly wages between 60 and 70 rupees.*

Wages in rupees		*No. of persons in thousands*
Below	*40*	*250*
40	*60*	*120*
60	*80*	*100*
80	*100*	*70*
100	*120*	*50*

SOLUTION. Interpolation of the number of persons earning wages below rupees.

Wages in Rupees per week (n)	No. of persons in thousands (y)	Differences First Δ^1	Second Δ^2	Third Δ^3	Fourth Δ^4
Below 40 x_0	250 y_0				
		120 Δ^1_0			
" 60 x_1	370 y_1		—20 Δ^2_0		
		100 Δ^1_1		—10 Δ^3_0	
" 80 x_2	470 y_2		—30 Δ^2_1		20 Δ^4_0
		70 Δ^1_2		+10 Δ^3_1	
" 100 x_3	540 y_3		—20 Δ^2_2		
		50 Δ^1_3			
" 120 x_4	590 y_4				

$$x=\frac{70-40}{80-60}=\frac{30}{20}=1.5$$

Substituting the above values in the Newton's formula of interpolation.

$$yx=y_3+x\Delta^1_0+\frac{x(x-1)}{1\times2}\Delta^2_0+\frac{x(x-1)(x-2)}{1\times2\times3}\Delta^3_0$$
$$+\frac{x(x-1)(x-2)(x-3)}{1\times2\times3}\Delta^4_0$$

Where yx represents the thing to be interpolated, that is, the number of persons earning wages below 70 rupees in this case.

We get,

$$yx=250+(1.5\times120)+\frac{1.5\,(1.5-1)\times-20}{1\times2}+\frac{1.5\,(1.5-1)\,(1.5-2)\times-10}{1\times2\times3}$$
$$+\frac{1.5\,(1.5-1)\,(1.5-2)\,(1.5-3)\times20}{1\times2\times3\times4}$$

Or $yx=250+180-7.5+.625+.46875=423.6$ thousand persons

Thus, the number of persons earning weekly wages below 70 rupees is 423.6 thousands,

and the number of persons earning wages below 60 rupees is 370 thousands.

Therefore the number of persons earning wages between 60 and 70 rupees is (423.6—370) *i.e.* 53.6 thousands.

PROBLEM 8.14. *Estimate the number of persons having incomes between 1000 and 1500 in the table given below in the groups A and B.*

Income in Rs.	*No. of persons Group A*	*No. of persons Group B*
Below 500	*6000*	*5000*
500—1000	*4250*	*4500*
1000—2000	*3600*	*4800*
2000—3000	*1500*	*2200*
3000—4000	*650*	*1500*

SOLUTION. Estimation of the number of persons having income below 1500 rupees in Groups A and B

Income in Rupees	No. of persons (cumulative frequency)	Group A			No. of persons (cumulative frequency)	Group B		
		Differences				Differences		
(x)	(y)	First Δ^1	Second Δ^2	Third Δ^3	(y)	First Δ^1	Second Δ^2	Third Δ^3
Below 1000	10250 y_0				9500 y_0			
		3600 Δ^1_0				4800 Δ^1_0		
,, 2000	13850 y_1		−2100 Δ^2_0		14300 y_1		−2600 Δ^2_0	
		1500 Δ^1_1		1250 Δ^3_0		2200 Δ^1_1		1900 Δ^3_0
,, 300	15350 y_2		− 850 Δ^2_1		16500 y_2		— 700 Δ^2_1	
		650 Δ^1_2				1500 Δ^1_2		
,, 4000	16000 y_3				18000 y_3			

$$x_a = x_b = \frac{\text{Income of interpolation—Income of origin}}{\text{Differences between two adjoining incomes}} = \frac{1500-1000}{2000-1000} = .5$$

Substituting the values in Newton's formula

$$yx = y_0 + x\Delta^1{}_0 + \frac{x(x-1)}{1\times 2}\Delta^2{}_0 + \frac{x(x-1)(x-2)}{1\times 2\times 3}\Delta^3{}_0$$

Where *yx* represents the thing to be interpolated, that is, the number of persons having incomes below Rs. 1500 in this case.

We get,

For Group A.

$$yx = 10250 + (.5\times 3600) + \frac{.5\,(.5-1)\times -2100}{1\times 2} + \frac{.5\,(.5-1)\,(.5-2)\times 1250}{1\times 2\times 3}$$

$$= 10250 + 1800 + 262.5 + 78.125 = 12390.625 \text{ or } 12391$$

For Group B.

$$yx = 9500 + (.5\times 4800) + \frac{.5\,(.5-1)\times -2600}{1\times 2} + \frac{.5\,(.5-1)\,(.5-2)\times 1900}{1\times 2\times 3}$$

$$= 9500 + 2400 + 325 + 118.75 = 12343.75 \text{ or } 12344$$

Thus, the number of persons having incomes below Rs. 1500 is 12391 in Group A and 12344 in Group B.

And the number of persons having incomes below Rs. 1000 is 10250 in Group A and 9500 in Group B.

Therefore the number of persons having income between Rs. 1000 and Rs. 1500 is 2141 in Group A and 2844 in Group B.

PROBLEM 8.15. *Find out by interpolation from the following data the number of workers earning Rs. 24 or more but less than Rs. 25 :*

Earning less than *Rs.*		*Number of workers*
20	..	*296*
25	..	*599*
30	..	*804*
35	..	*918*
40	..	*966*

SOLUTION. Estimation of the number of workers earning less than Rs. 24 by Newton's Method.

Earnings in Rs. (*x*)		Number of Workers (*y*)		Differences First Δ^1		Second Δ^2		Third Δ^3		Fourth Δ^4	
Less than 20	x_0	296	y_0								
				303	$\Delta^1{}_0$						
,, ,, 25	x_1	599	y_1			−98	$\Delta^2{}_0$				
				205	$\Delta^1{}_1$			7	$\Delta^3{}_0$		
,, ,, 30	x_2	804	y_2			−91	$\Delta^2{}_1$			18	$\Delta^4{}_0$
				114	$\Delta^1{}_2$			25	$\Delta^3{}_1$		
,, ,, 35	x_3	918	y_3			−56	$\Delta^2{}_2$				
				48	$\Delta^1{}_3$						
,, ,, 40	x_4	966	y_4								

$$x = \frac{\text{Earning of interpolation—Earning of origin}}{\text{Difference between two adjoining earnings}}$$

$$= \frac{24-20}{25-20} = \frac{4}{5} = .8$$

Substituting the above values in the Newton's formula of interpolation.

$$yx = y_0 + x\Delta^1{}_0 + \frac{x(x-1)}{1\times 2}\,\Delta^2{}_0 + \frac{x(x-1)\,(x-2)}{1\times 2\times 3}\,\Delta^3{}_0 + \frac{x(x-1)\,(x-2)\,(x-3)}{1\times 2\times 3\times 4}\,\Delta^4{}_0\,;$$

Where yx represents the thing to be interpolated, that is, the number of workers earning less than Rs. 24, in this case.

We get,

$$yx = 296 + (.8\times 303) + \frac{.8(.8-1)\times -98}{1\times 2} + \frac{.8(.8-1)\,(.8-2)\times 7}{1\times 2\times 3} + \frac{.8\,(8-1)\,(.8-2)\,(.8-3)\times 18}{1\times 2\times 3\times 4}$$

$= 296 + 242.4 + 7.84 + .224 - .3168 = 546.1472$ or

$= 546.$

Thus, the number of workers earning less than Rs. 24 is 546
and ,, ,, ,, ,, ,, Rs. 25 is 599

Therefore, the number of workers earning Rs. 24 or more but less than Rs. 25 is $(599-546) = 53$.

PROBLEM 8.16. *The following table relates to income earned per day by a certain number of workers in a big manufacturing concern.*

Earnings per day in rupees	*Number of workers*
Up to 10	*50*
20	*150*
30	*300*
40	*500*
50	*700*
60	*800*

It is required to find out the number of workers (a) falling with in Rs. 25 - 35 earning and (b) earning above Rs. 42.

SOLUTION. Interpolation of the number of workers whose earnings are up to Rs. 25, Rs. 35 and Rs. 42 per day.

Earnings per day in rupees (x)		Number of Workers (y)	Differences First Δ^1	Second Δ^2	Third Δ^3	Fourth Δ^4	Fifth Δ^5
Upto 10	x_0	50 y_0					
			100 $\Delta^1{}_0$				
,, 20	x_1	150 y_1		50 $\Delta^3{}_0$			
			150 $\Delta^1{}_1$		0 $\Delta^3{}_0$		
,, 30	x_2	300 y_2		50 $\Delta^2{}_1$		—50 $\Delta^4{}_0$	
			200 $\Delta^1{}_2$		—50 $\Delta^3{}_1$		0 $\Delta^5{}_0$
,, 40	x_3	500 y_3		0 $\Delta^2{}_2$		—50 $\Delta^4{}_1$	
			200 $\Delta^1{}_3$		—100 $\Delta^3{}_2$		
,, 50	x_4	700 y_4		—100 $\Delta^2{}_3$			
			100 $\Delta^1{}_4$				
,, 60	x_5	800 y_5					

$$x = \frac{\text{Earning of interpolation}-\text{Earning of origin}}{\text{Difference between two adjoining earnings}}$$

Substituting the values, we get,

$$x_{25} = \frac{25-10}{30-20} = \frac{15}{10} = 1.5$$

$$x_{35} = \frac{35-10}{40-30} = \frac{25}{10} = 2.5$$

$$x_{42}=\frac{42-10}{50-40}=\frac{32}{10}=3.2$$

Applying to the above data, Newton's formula of interpolation.

$$yx=y_0+x\Delta^1{}_0+\frac{x(x-1)}{1\times2}\Delta^2{}_0+\frac{x(x-1)(x-2)}{1\times2\times3}\Delta^3{}_0$$
$$+\frac{x(x-1)(x-2)(x-3)}{1\times2\times3\times4}\Delta^4{}_0+\frac{x(-1)(x-2)(x-3)(x-4)}{1\times2\times3\times4\times5}\Delta^5{}_0$$

Where yx represents the things to be interpolated.

We, get,

(1) *Number of workers whose earnings are upto Rs. 25.*

$$yx_{25}=50+(1.5\times100)+\frac{1.5\,(1.5-1)\times50}{1\times2}+\frac{15(1.5-1)\,(1.5-2)\times0}{1\times2\times3}$$
$$+\frac{1.5(1.5-1)\,(1.5-2)(1.5-3)\times50}{1\times2\times3\times4}$$
$$+\frac{1.5\,(1.5-1)\,(1.5-2)\,(1.5-3)\,(1.5-4)\times0}{1\times2\times3\times4\times5}$$
$$=50+150+18.75-0-1.1719-0=217.5781 \text{ or } 218$$

(2) *Number of workers whose earnings are upto Rs. 35.*

$$=50+(2.5\times100)+\frac{2.5\,(2.5-1)\times50}{1\times2}+\frac{2.5\,(2.5-1)\,(2.5-2)\times0}{1\times2\times3}$$
$$+\frac{2.5\,(2.5-1)\,(2.5-2)\,(2.5-3)\times-50}{1\times2\times3\times4}$$
$$+\frac{2.5\,(2.5-1)\,(2.5-2)\,(2.5-3)\,(2.5-4)\times0}{1\times2\times3\times4\times5}$$
$$=50+250+93.75+0+1.9531+0=365.7031 \text{ or } 396$$

(3) *Number of workers whose earnings are upto Rs. 42.*

$$yx_{42}=50+(3.2-100)+\frac{3.2\,(3.2-1)\times50}{1\times2}+\frac{3.2\,(3.2-1)\,(3.2-2)\times0}{1\times2\times3}$$
$$+\frac{3.2\,(3.2-1)\,(3.2-2)\,(3.2-3)\times-50}{1\times2\times3\times4}$$
$$+\frac{3.2\,(3.2-1)\,(3.2-2)\,(3.2-3)\,(3.2-4)\times0}{1\times2\times3\times4\times5}$$
$$=50+320+176+0-3.52-0=542.48 \text{ or } 543$$

Thus, the number of workers earning upto Rs. 25 is 218
and „ „ „ „ Rs. 35 is 396

Therefore the number of workers falling within Rs. 25-35 earning group is 396—218 = 178

Similarly the number of workers earning upto Rs. 42 is 543
and the total number of workers is 800

Therefore the number of workers earning above Rs. 42 is 800—543 = 257.

PROBLEM 8.17. *The following are the marks obtained by 492 candidates in a certain examination :—*

Not more than 40 marks 212 candidates
Not more than 45 marks 296 candidates
Not more than 50 marks 368 candidates
Not more than 55 marks 429 candidates
Not more than 60 marks 460 candidates
Not more than 65 marks 481 candidates
Not more than 70 marks 490 candidates
Not more than 75 marks 492 candidates

Find out the number of candidates who secured more than 42 but not more than 45 marks.

$$x = \frac{\text{Marks of interpolation—Marks of origin}}{\text{Difference between two adjoining marks}} = \frac{42-40}{45-40} = .4$$

Substituting the above values in Newton's formula of interpolation,

$$yx = y_0 + x\triangle^1_0 + \frac{x(x-1)}{1\times2}\triangle^2_0 + \frac{x(x-1)(x-2)}{1\times2\times3}\triangle^3_0 + \frac{x(x-1)(x-2)(x-3)}{1\times2\times3\times4}\triangle^4_0 + \frac{x(x-1)(x-2)(x-3)(x-4)}{1\times2\times3\times4\times5}\triangle^5_0 + \frac{x(x-1)(x-2)(x-3)(x-4)(x-5)}{1\times2\times3\times4\times5\times6}\triangle^6_0 + \frac{x(x-1)(x-2)(x-3)(x-4)(x-5)(x-6)}{1\times2\times3\times4\times5\times6\times7}\triangle^7_0$$

Where *yx* represents the thing to be interpolated, that is, the number of candidates securing not more than 42 marks in this case.

$$yx_{42} = 212 + (.4\times84) + \frac{.4(.4-1)\times-12}{1\times2} + \frac{.4(.4-1)(.4-2)\times1}{1\times2\times3} + \frac{4(.4-1)(.4-2)(.4-3)\times-20}{1\times2\times3\times4} + \frac{.4(.4-1)(.4-2)(.4-3)(.4-4)\times59}{1\times2\times3\times4\times5} + \frac{.4(.4-1)(.4-2)(.4-3)(.4-4)(.4-5)\times-120}{1\times2\times3\times4\times5\times6} + \frac{.4(.4-1)(.4-2)(.4-3)(.4-4)(.4-5)(.4-6)\times210}{1\times2\times3\times4\times5\times6\times7}$$

$$\text{Or } yx = 212 + 33.6 + 1.44 + .064 + .832 + 1.7672 + 2.756 + 3.858$$
$$= 256.3172 = 256$$

Thus, the number of candidates getting not more than 42 marks is 256.
and the number of candidates getting not more than 45 marks is 296.

Therefore the number of candidates who secured more than 42 but not more than 45 marks is 296—256 = 40.

PROBLEM 8.18. *The following are the number of deaths in four successive ten-year age groups. Calculate the number of deaths at 45—50 and 50—55.*

Age-group	*Deaths.*
upto 25	*13,229*
,, 35	*18,139*
,, 45	*24,225*
,, 55	*31,496*

SOLUTION. Interpolation of number of deaths upto the age of 50.

Age-group (x)		Deaths (y)		Differences					
				First $\triangle^1$		Second $\triangle^2$		Third $\triangle^3$	
Upto 25	x_0	13,229	y_0						
				+4,910	$\triangle^1_0$				
,, 35	x_1	18,139	y_1			+1,176	$\triangle^2_0$		
				+6,086	$\triangle^1_1$			+9	$\triangle^3_0$
,, 45	x_2	24,225	y_2			+1,185	$\triangle^2_1$		
				+7,271	$\triangle^1_2$				
,, 55	x_3	31,496	y_3						

$$x = \frac{\text{Age of interpolation—Age of origin}}{\text{Difference between two adjoining ages}}$$

Substituting the values, we get,

$$x50 = \frac{50—25}{55—45} = \frac{25}{10} = 2.5$$

Applying to the above data, Newton's formula of interpolation.

$$yx = y_0 + x\triangle^1{}_0 + \frac{x(x—1)}{1\times2}\triangle^2{}_0 + \frac{x(x—1)(x—2)}{1\times2\times3}\triangle^3{}_0$$

Where yx represents the thing to be interpolated.

We get,

The number of deaths at the age 50.

$$yx50 = 13{,}229 + (2.5\times4{,}910) + \frac{2.5\,(2.5—1)\times1{,}176}{1\times2} + \frac{2.5\,(2.5—1)\,(2{\cdot}5—2)\times9}{1\times2\times3}$$

$$= 13{,}229 + 12{,}275 + 2{,}205 + 2.8125$$
$$= 27{,}711.8125 \quad = 27{,}712 \text{ deaths}$$

Thus, the number of deaths at age 50 is 27,712
and ,, ,, ,, 45 is 24,225

Therefore the number of deaths between 45—50 age group is 27,712—24,225 = 3,487

Similarly the number of deaths upto 55 is 31,496
and ,, ,, ,, 50 is 27,712

Therefore, the number of deaths between 50 to 55 age group is 31,496—27,712 = 3,784.

PROBLEM 8.19. *The following are the amounts of income tax paid by 600 businessmen of certain district of Uttar Pradesh in the year 1970 :—*

More than Rupees	*500*	..	*600*
More than Rupees	*1000*	..	*550*
More than Rupees	*1500*	..	*425*
More than Rupees	*2000*	..	*275*
More than Rupees	*2500*	..	*100*
More than Rupees	*3000*	..	*25*

Find out the number of businessmen who paid more than Rs. 1,200 but not more than Rs. 2,400 as income tax.

SOLUTION. Interpolation of the number of businessmen who paid more than Rs. 1,200 but not more than Rs. 2,400 as income tax.

Amount of income tax in rupees	Number of businessmen		Difference									
			First $\triangle^1$		Second $\triangle^2$		Third $\triangle^3$		Fourth $\triangle^4$		Fifth $\triangle^5$	
More than Rs. 500	600	y_0										
			— 50	$\triangle^1{}_0$								
,, ,, 1000	550	y_1			— 75	$\triangle^2{}_0$						
			—125	$\triangle^1{}_1$			+ 50	$\triangle^3{}_0$				
,, ,, 1500	425	y_2			— 25	$\triangle^2{}_1$			— 50	$\triangle^4{}_0$		
			—150	$\triangle^1{}_2$			0	$\triangle^3{}_1$			+175	$\triangle^5{}_0$
,, ,, 2000	275	y_3			— 25	$\triangle^2{}_2$			+125	$\triangle^4{}_1$		
			—175	$\triangle^1{}_3$			+125	$\triangle^3{}_2$				
,, ,, 2500	100	y_4			+100	$\triangle^2{}_3$						
			— 75	$\triangle^1{}_4$								
,, ,, 3000	25	y_5										

$$x = \frac{\text{Income tax of interpolation—Income Tax of Origin}}{\text{Difference between two adjoining amounts of income tax}}$$

Substituting the values, we get,

$$x_{1200} = \frac{1200-500}{1500-1000} = \frac{700}{500} = 1.4$$

$$x_{2400} = \frac{2400-500}{2500-2000} = \frac{1900}{500} = 3.8$$

Applying to the above data Newton's formula of interpolation.

$$yx = y_0 + x\Delta^1_0 + \frac{x(x-1)}{1\times2}\Delta^2_0 + \frac{x(x-1)(x-2)}{1\times2\times3}\Delta^3_0$$

$$+\frac{x(x-1)(x-2)(x-3)}{1\times2\times3\times4}\Delta^4_0 + \frac{x(x-1)(x-2)(x-3)(x-4)}{1\times2\times3\times4\times5}\Delta^5_0$$

Where yx represents the thing to be interpolated,

We get,

1. *Number of businessmen who paid more than Rs. 1200 as income tax.*

$$yx_{1200} = 600 + (1.4\times-50) + \frac{1.4\,(1.4-1)\times-75}{1\times2}$$

$$+\frac{1.4\,(1.4-1)\,(1.4-2)\times50}{1\times2\times3} + \frac{1.4\,(1.4-1)\,(1.4-2)\,(1.4-3)\times-50}{1\times2\times3\times4}$$

$$+\frac{1.4\,(1.4-1)\,(1.4-2)\,(1.4-3)\,(1.4-4)\times175}{1\times2\times3\times4\times5}$$

or $yx_{1200} = 600-70-21-2.8-1.12-2.0384 = 503.0416$

or $yx_{1200} = 503$

2. *Number of businessmen who paid more than Rs. 2400 as income tax.*

$$yx_{2400} = 600 + (3.8\times-50) + \frac{3.8(3.8-1)\times-75}{1\times2}$$

$$+\frac{3.8\,(3.8-1)\,(3.8-2)\times50}{1\times2\times3} + \frac{3.8\,(3.8-1)\,(3.8-2)(3.8-3)\times-50}{1\times2\times3\times4}$$

$$+\frac{3.8\,(3.8-1)\,(3.8-2)\,(3.8-3)\,(3.8-4)\times175}{1\times2\times3\times4\times5}$$

or $yx_{2400} = 600-190-399+159.6-31.92-4.4688 = 134.212$

or $yx_{2400} = 134.$

Thus the number of businessmen who paid more than Rs. 1200 as income-tax is 503.

and the number of businessmen who paid more than Rs. 2400 as income-tax is 134.

Therefore, the number of businessmen who paid more than Rs. 1200 but not more than Rs. 400 as income-tax is (503-134) = 369.

PROBLEM 8.20. *What are the assumptions on which the method of interpolation is based ? Find out Newton's formula for interpolation in case of equal intervals.*

The following are the marks obtained by 492 candidates in a certain examination :

Not more than 40 marks, 210 candidates
Not more than 45 marks, 253 candidates
Not more than 50 marks, 307 candidates
Not more than 55 marks, 381 candidates
Not more than 60 marks, 413 candidates
Not more than 65 marks 492 candidates

Find out the number of candidates (a) who secured more than 48 but not more than 50 marks. (b) less than 48 but not less than 45 marks.

SOLUTION. Interpolation of the number of candidates securing more than 48 marks.

Marks not more than	Number of Candi-dates		Differences									
			First Δ^1		Second Δ^2		Third Δ^3		Fourth Δ^4		Fifth Δ^5	
40	210	y_0										
			+43	Δ^1_0								
45	253	y_1			+11	Δ^2_0						
			+54	Δ^1_1			+ 9	Δ^3_0				
50	307	y_2			+20	Δ^2_1			−71	Δ^4_0		
			+74	Δ^1_2			−62	Δ^3_1			+222	Δ^5_0
55	381	y_3			−42	Δ^2_2			+151	Δ^4_1		
			+32	Δ^1_3			+89	Δ^3_2				
60	413	y_4			+47	Δ^2_3						
			+79	Δ^1_4								
65	492	y_5										

$$x=\frac{\text{Marks of interpolation—Marks of origin}}{\text{Difference between two adjoining marks}}=\frac{48-40}{50-45}=\frac{8}{5}=1.6$$

Substituting the above values in Newton's formula of interpolation,

$$y_x=y_0+x\Delta^1_0+\frac{x(x-1)}{1\times2}\Delta^2_0+\frac{x(x-1)(x-2)}{1\times2\times3}\Delta^3_0$$
$$+\frac{x(x-1)(x-2)(x-3)}{1\times2\times3\times4}\Delta^4_0+\frac{x(x-1)(x-2)(x-3)(x-4)}{1\times2\times3\times4\times5}\Delta^5_0$$

Where yx represents the thing to be interpolated, that is, the number of candidates securing not more than 48 marks in this case.

$$yx_{48}=210+(1.6\times43)+\frac{1.6\,(1.6-1)\times11}{1\times2}+\frac{1.6\,(1.6-1)\,(1.6-2)\times9}{1\times2\times3}$$
$$+\frac{1.6\,(1.6-1)\,(1.6-2)\,(1.6-3)\times-71}{1\times2\times3\times4}$$
$$+\frac{1.6\,(1.6-1)\,(1.6-2)\,(1.6-3)\,(1.6-4)\times222}{1\times2\times3\times4\times5}$$

or $yx_{48}=210+68.8+5.28-.576-1.5904-2.386944=284.08-4.553344$
$279.526656=280$ candidates.

(*a*) Thus the number of the candidates getting not more than 48 marks is 280.

and the number of candidates getting more than 50 marks is 307.

Therefore, the number of candidates who secured more than 48 marks but not more than 50 marks is 307—280 = 27.

(*b*) The number of candidates getting not more than 48 marks is 280. and the number of candidates getting not more than 45 marks is 253.

Therefore, the number of candidates who secured less than 48 but not less than 45 is 280-253 = 27.

PROBLEM 8.21 *From the following table, find the number of students who obtained less than 45 marks :—*

Marks	*No. of students*
30—40	*31*
40—50	*42*
50—60	*51*
60—70	*35*
70—80	*31*

SOLUTION. Interpolation of the number of students who obtained less than 45 marks.

Marks x		No. of Students		Differences							
				First $\triangle^1$		Second $\triangle^2$		Third $\triangle^3$		Fourth $\triangle^4$	
Less than											
40	x_0	31	y_0	42	$\triangle^1_0$	+ 9	$\triangle^2_0$				
50	x_1	73	y_1	51	$\triangle^1_1$	—16	$\triangle^2_1$	—25	$\triangle^3_0$	+37	$\triangle^4_0$
60	x_2	124	y_2	35	$\triangle^1_2$	— 4	$\triangle^2_2$	+12	$\triangle^3_1$		
70	x_3	159	y_3	31	$\triangle^1_3$						
80	x_4	190	y_4								

$$x = \frac{\text{Marks of interpolation—Marks of origin}}{\text{Difference between two adjoining marks}}$$

$$= \frac{45—40}{50—40} = \frac{5}{10} = .5$$

Applying to the above data, Newton's formula

$$yx = y_0 + x\triangle^1_0 + \frac{x(x—1)}{1 \times 2}\triangle^2_0 + \frac{x(x—1)(x—2)}{1 \times 2 \times 3}\triangle^3_0 + \frac{x(x—1)(x—2)(x—3)}{1 \times 2 \times 3 \times 4}\triangle^4_0$$

We get,

$$yx_{45} = 31 + (.5 \times 42) + \frac{.5(.5—1) \times 9}{1 \times 2} + \frac{.5(.5—1)(.5—2) \times —25}{1 \times 2 \times 3} + \frac{.5(.5—1)(.5—2)(.5—3) \times 37}{1 \times 2 \times 3 \times 4}$$

$$= 31 + 21.0—1.125—1.562—1.445 = 47.968 = 48$$

Thus the number of students who obtained less than 45 marks is 48.

PROBLEM 8.22. *Extrapolate population of a town for 1976 from the following data about its population during the previous four censuses :—*

Census Year	*Population in thousands*
1941	*473*
1951	*468*
1961	*454*
1971	*484*

SOLUTION. Extrapolation of the population for 1976.

Year		Population (000)		Differences					
				First $\triangle^1$		Second $\triangle^2$		Third $\triangle^3$	
1941	x_0	473	y_0						
1951	x_1	468	y_1	— 5	$\triangle^1_0$				
1961	x_2	454	y_2	—14	$\triangle^1_1$	—	$\triangle^2_0$		
1971	x_3	484	y_3	+30	$\triangle^1_2$	+44	$\triangle^2_1$	+53	$\triangle^3_0$

$$x = \frac{\text{Year of extrapolation—year of origin}}{\text{Difference between two adjoining year}}$$

$$= \frac{1976-1941}{1951-1941} = \frac{35}{10} = 3.5$$

Applying to the above data. Newton's formula

$$yx = y_0 + x\Delta^1_0 + \frac{x(x-1)}{1\times 2}\Delta^2_0 + \frac{x(x-1)(x-2)}{1\times 2\times 3}\Delta^3_0$$

We get,

$$yx = 473 + (3.5\times -5) + \frac{3.5(3.5-1)\times -9}{1\times 2} + \frac{3.5(3.5-1)(3.5-2)\times 53}{1\times 2\times 3}$$

$$= 473 - 17.5 - 39.375 + 115.937 = 532.062$$

Thus the extrapolated figure of the population for the year 1976 is 532.06 (thousands)

PROBLEM 8.23 (*a*) *By constructing a difference table, find the 7th term a well as the general term of the sequence :*

0, 0, 2, 6, 12, 20, ...30.

(*b*) *Given*

$\sin 45^0 = 0.7071,$
$\sin 50^0 = 0.7660,$
$\sin 55^0 = 0.8192,$
$\sin 60^0 = 0.8660.$

Find sin 52^0, *by using any method of interpolation.*

SOLUTION.

(*a*)

		Δ^1	Δ^2	Δ^3
0	y_0			
0	y_1	0 Δ^1_0		
2	y_2	2 Δ^1_1	2 Δ^2_0	0 Δ^3_0
6	y_3	4 Δ^1_2	2 Δ^2_1	0 Δ^3_1
12	y_4	6 Δ^1_3	2 Δ^2_2	0 Δ^3_2
20	y_5	8 Δ^1_4	2 Δ^2_3	
~				
...				
30				

The value of the 7th term or

$y_6 = y_5 + \Delta^1_5 = y_5 + \Delta^1_4 + \Delta^2_4$

Since the second differences are all equal $\Delta^2_4 = 2$

$y_6 = 20 + 8 + 2 = 30$

Ordinarily the general term of sequence in Newton's formula is

$$yx - y_0 + x\Delta^1_0 + \frac{x(x-1)}{1\times 2}\Delta^2_0 + \ldots$$

In the above case $y_0=0$, and $\triangle^1_0$ is also 0. The value of $\triangle^2_0=2$ and $\triangle^3_0=0$. Therefore the general term of the sequence would be

$$yx=0+0+\frac{x(x-1)}{2}\times 2 \text{ or } yx=x(x-1)$$

(*b*) Interpolation of the sin 52° by Newton's method

			Differences					
Sin	Value of sin		First $\triangle^1$		Second $\triangle^2$		Third $\triangle^3$	
45	0.7071	y_0						
			+0.0589	$\triangle^1_0$				
50	0.7660	y_1			-0.0057	$\triangle^2_0$		
			+0.0532	$\triangle^1_1$			-0.0007	$\triangle^3_0$
55	0.8192	y_2			-0.0064	$\triangle^2_1$		
			+0.0468	$\triangle^1_2$				
60	0.8660	y_3						

$$x=\frac{52-45}{55-50}=\frac{7}{5}=1.4$$

Substituting the above values in the Newton's formula of interpolation.

$$yx=y_0+x\triangle^1_0+\frac{x(x-45)}{1\times 2}\triangle^2_0+\frac{x(x-1)(x-2)}{1\times 2\times 3}\triangle^3_0$$

Where yx represents the thing to be interpolated, that is, the value of sin 52°

$$yx=0.7071+(1.4\times 0.0589)+\frac{1.4(1.4-1)\times -.0057}{1\times 2}+\frac{1.4(1.4-1)(1.4-2)\times -.0007}{1\times 2\times 3}$$

or $yx=0.7071+.08246-.001596+.0000392$

or $yx=.7895992-.001596$ or $yx=.7880032$

Thus the value of sin 52° is .7880.

NEWTON'S GAUSS AND STERLING FORMULA

Newton's Gauss :—The formula should be applied when the missing figure is in the middle of the table. The formula is,

$$yx=y_0+x\triangle^1y_0+\frac{x(x-1)}{1\times 2}\triangle^2y_{-1}+\frac{(x-1)\,x\,(x-1)}{1\times 2\times 3}\triangle^3y_{-1}+\frac{(x-1)\,x\,(x-1)(x-2)}{1\times 2\times 3\times 4}\triangle^4y_{-2}+\ldots\ldots$$

PROBLEM 8.24. *Estimate the value of y if $x=3.75$ from the following table* :—

x	*y*
2.5	*24.145*
3.0	*22.043*
3.5	*20.225*
4.0	*18.644*
4.5	*17.262*
5.0	*16.047*

SOLUTION. In this method y_0 is generally the value of the item immediately preceding the interpolated value :—

x	y		Differences First Δ^1		Second Δ^2		Third Δ^3		Fourth Δ^4	
2.5	24.145	y_{-2}								
			-2.102	$\Delta^1 y_{-2}$						
3.0	22.043	y_{-1}			$+.284$	$\Delta^2 y_{-2}$				
			-1.818	$\Delta^1 y_{-1}$			$-.047$	$\Delta^3 y_{-2}$		
3.5	20.225	y^0			$+.237$	$\Delta^2 y_{-1}$			$+.009$	$\Delta^4 y_{-2}$
			-1.581	$\Delta^1 y_0$			-038	$\Delta^3 y_{-1}$		
4.0	18.644	y^1			$+.199$	$\Delta^2 y_0$			$+.006$	$\Delta^4 y_{-1}$
			-1.382	$\Delta^1 y_1$			$-.032$	$\Delta^3 y_0$		
4.5	17.262	y^2			$+.167$	$\Delta^2 y_1$				
			-1.215	$\Delta^1 y_2$						
5.0	16.047	y^3								

Newton's Gauss formula :

$$yx = yo + x\Delta^1 y_0 + \frac{x(x-1)}{1\times 2}\Delta^1 y_{-1} + \frac{(x+1)\,x\,(x-1)}{1\times 2\times 3}\Delta^2 y_{-2}$$

$$+\frac{(x+1)\,x\,(x-1)\,(x-2)}{1\times 2\times 3\times 4}\Delta^4 y_{-2}$$

$$x + \frac{3.75-3.5}{4.0-3.5} = \frac{.25}{5} = .5$$

Substituting the value we get,

$$yx = 20.225 + (.5\times -1.581) + \left\{\frac{.5\,(.5-1)\times .237\,)}{2}\right\}$$

$$+\left\{\frac{(.5+1)\,.5\,(.5-1)\times -.038}{6}\right\}$$

$$+\left\{\frac{(.5+1)\,.5\,(.5-1)\,(.5-2)\times .009}{24}\right\}$$

$$= 20.225 - .7905 - .029625 + .0023750 + .0002109 = 19.4074609$$

19.407 approx.

Thus the value of y for x 3.75 = 19.407.

PROBLEM 8.25. *The following table gives the sales of a concern for the last few years. Estimate the sales for the year 1971.*

Year	*Sales (Rs. 000)*
1966	*40*
1968	*43*
1970	*48*
1972	*52*
1974	*57*

SOLUTION. Interpolation of the sales for the year 1971.

Year	Sales (Rs. 000)		Differences							
x	(y)		First $\triangle^1$		Second $\triangle^2$		Third $\triangle^3$		Fourth $\triangle^4$	
1966	40	y_{-2}								
			+3	$\triangle^1 y_{-2}$						
1968	43	y_{-1}			+2	$\triangle^2 y_{-2}$				
			+5	$\triangle^1 y_{-1}$			—3	$\triangle^3 y_{-2}$		
1970	48	y_0			—1	$\triangle^2 y_{-1}$			$\triangle 5$	$\triangle^4 y_{-2}$
			+4	$\triangle^1 y_0$			+2	$\triangle^3 y_{-1}$		
1972	52	y_1			+1	$\triangle^1 y_0$				
			+5	$\triangle^1 y_1$						
1974	57	y_2								

Newton's Gauss formula is :

$$yx = yo + x\triangle^1 y_0 + \frac{x(x-1)}{1\times 2}\triangle^2 y_{-1} + \frac{(x+1)\,x\,(x-1)}{1\times 2\times 3}\triangle^3 y_{-1}$$
$$+\frac{(x+1)\,x\,(x-1)\,(x-2)}{1\times 2\times 3\times 4}\triangle^4 y_{-2}$$

$$x = \frac{(1971-1970)}{(1972-1970)} = .5$$

$$yx = 48 + (.5\times 4) + \frac{.5(.5-1)\times 1}{2} + \frac{(1.5\times .5\times -.5\times 2)}{6}$$
$$+\frac{(1.5\times .5\times -.5\times -1.5\times 5)}{24}$$

$$= 48 + 2 + .125 - .125 + .117 = 50.117.$$

Thus the estimated sales for the year 1951 are of 50.117 thousand rupees.

STERLING'S FORMULA

This formula is also used when the missing figure is in the middle of the table. The formula

$$yx = y_0 + x\,\frac{\triangle^1 y_{-1} + \triangle^1 y_0}{2} + \frac{x^2}{2}\triangle^2 y_{-1}$$
$$+\frac{x(x^2-1^2)}{6}\times\frac{\triangle^3 y_2 + \triangle^3 y_{-1}}{2} + \frac{x^2}{24}(x^2-1^2)\,\triangle^4 y_{-2}$$

PROBLEM 8.26. *Using Sterling's formula interpolate the value of y when* $x = 9$.

x	y
4	40
6	43
8	48
10	52
12	57

SOLUTION. Interpolation of the value of y when $x=9$.

x	y		Differences							
			First Δ^1		Second Δ^2		Third Δ^3		Fourth Δ^4	
4	40	y^{-2}								
			+3	$\Delta^1 y_{-2}$						
6	43	y^{-1}			+2	$\Delta^2 y_{-2}$				
			+5	$\Delta^1 y_{-1}$			—3	$\Delta^3 y_{-2}$		
8	48	y^0			—1	$\Delta^2 y_{-1}$			+5	$\Delta^4 y_{-2}$
			+4	$\Delta^1 y_0$			+2	$\Delta^3 y_{-1}$		
10	52	y^1			+1	$\Delta^2 y_0$				
			+5	$\Delta^1 y_1$						
12	57	y^2								

Value of $x = \frac{9-8}{8-6} = .5$

Sterling's formula is

$$yx = y_{)} + x \frac{\Delta^1 y_{-1} + \Delta^1 y_0}{2} + \frac{x^2}{2} \Delta^2 y_{-1} + \frac{x(x^2-1^2)}{6} \times \frac{\Delta^3 y_{-2} + {}^3 y_{-1}}{2} + \frac{x^2}{24}(x^2-1^2)\Delta^4 y_{-2} \ldots\ldots$$

Substituting the values, we get

$$yx = 48 + .5 \frac{5+4}{2} + \frac{.25}{2} \times -1 + \frac{.5(.25-1)}{6} \times \frac{-3+2}{2} + \frac{.25}{24}(.25-1)5$$

$= 48 + 2.25 - .125 + 03115 - 039 = 50.117.$

PROBLEM 8.27. *If lx represents the numbers living at age x in a life table, find lx for values of x = 35. Use Sterling's formula :*

$l_{20} = 512$; $l_{30} = 439$, $l_{40} = 346$, $l_{50} = 243$

SOLUTION. Interpolation of the value of l_{35}.

x	y		Differences					
			First Δ^1		Second Δ^2		Third Δ^3	
20	512	y^{-1}						
			— 73	$\Delta^1 y_{-1}$				
30	439	y_0			—20	$\Delta^2 y_{-1}$		
			— 93	$\Delta^1 y_0$			+10	$\Delta^3 y_{-1}$
40	346	y^1			—10	$\Delta^2 y_0$		
			—103	$\Delta^1 y_1$				
50	243	y^2						

Value of $x = \frac{35-30}{30-20} = \frac{5}{10} = .5$

$$yx = y_0 + x \frac{\Delta^1 y_{-1} + \Delta^1 y_0}{2} + \frac{x^2}{2} \Delta^2 y_{-1} + \frac{x(x^2-1^2)}{6} \times \frac{\Delta^3 y_{-2} + \Delta^3 y_{-1}}{2}$$

$= 439 + .5 \frac{(-73)+(-93)}{2} + \frac{.25}{2} 20$

$= 439 - 41.5 - 2.5 = 395$

Thus the value of $l\,35 = 395$.

NEWTON'S GAUSS (BACKWARD) FORMULA

This formula of Newton's Gauss is used when the figure to be interpolated is at the end of the table.

The formula is

$$yx = y_0 - x\triangle^1 y_{-1} + \frac{(x+1)x}{2}\triangle^2 y_{-1} - \frac{(x+1)\,x\,(x-1)}{6}\triangle^3 y_{-2} + \frac{(x+1)\,x\,(x-1)\,(x-2)}{24}\triangle^4 y_{-2}$$

In this formula y_0 is the figure succeeding the missing figure ; and $x =$ The year or other unit succeeding the interpolation year—the interpolation year divided by difference of two adjoining years.

PROBLEM 8.28. *The following table gives the population of a town during the last 5 censuses. Estimate the population for the year 1956.*

Year	*Population (000)*
1921	*12*
1931	*15*
1941	*20*
1951	*27*
1961	*39*
1971	*52*

SOLUTION. Interpolation of the population for the year 1956.

Year x	Population (000)	y	Differences First $\triangle^1$		Second $\triangle^2$		Third $\triangle^3$		Fourth $\triangle^4$	
1921	12	y_{-4}								
			+3	$\triangle^1 y_{-4}$						
1931	15	y_{-3}			+2	$\triangle^2 y_{-4}$				
			+5	$\triangle^1 y_{-3}$			0	$\triangle^2 y_{-4}$		
1941	20	y_{-2}			+2	$\triangle^2 y_{-3}$			+3	$\triangle^4 y_{-4}$
			+7	$\triangle^1 y_{-2}$			+3	$\triangle^3 y_{-3}$		
1951	27	y_{-1}			+5	$\triangle^9 y_{-2}$			—7	$\triangle^4 y_{-3}$
			+12	$\triangle^1 y_{-1}$			—4	$\triangle^2 y_{-2}$		
1961	39	$y0$			+1	$\triangle^2 y_{-1}$				
			+13	$\triangle^1 y_{0}$						
1971	52	$y1$								

$$x = \frac{1961-1956}{1961-1951} = \frac{5}{10} = .5$$

$$x = y_0 - x\triangle^1 y_{-1} + \frac{(x+1)x}{1}\triangle^2 y_{-1} - \frac{(x+1)\,x\,(x-1)}{6}\triangle^3 y_{-2}$$

$$= 39-(.5\times 12)+\frac{(1.5\times .5\times 1)}{2}-\left(\frac{1.5\times .5-.5}{6}\times -4\right)$$

$$= 39-6+.375-.250 = 33.125 \text{ thousands.}$$

PROBLEM 8.29.

Year	*Sales (Rs. 000)*
1950	*25*
1955	*32*
1960	*40*
1965	*47*
1970	*55*
1975	*64*

From the above table estimate the sales in the year 1968. Use Newton's Gau (backward) formula.

SOLUTION. Interpolation of the sales of the year 1943.

Year	Sales (Rs. 000)		Differences							
			First		Second		Third		Fourth	
x	y		Δ^1		Δ^2		Δ^3		Δ^4	
1950	25	y-4								
			7	Δ^1y_{-4}						
1955	32	y-3			1	Δ^2y_{-4}				
			8	Δ^1y_{-3}						
1960	40	y-2			—1	Δ^2y_{-3}	—2	Δ^3y_{-4}		
			7	Δ^1y_{-2}					4	Δ^4y_{-4}
1965	47	y-1			1	Δ^2y_{-2}	+2	Δ^3y_{-3}		
			8	$\Delta^1y_{=1}$					—2	Δ^2y_{-3}
1970	55	y0			1	Δ^2v_{-1}	0	Δ^3y_{-2}		
			9	Δ^1y_0						
1975	64	y1								

$$x = \frac{1970-1968}{1970-1965} = \frac{2}{5} = .4.$$

$$yx = y_0 - x\Delta^1y_{-1} + \frac{(x+1)x}{2}\Delta^2y_{-1} - \frac{(x+1)\,x\,(x-1)}{6}\Delta^3y_{-2}$$

Substituting the values, we get,

$$yx = 55-(.4\times8)+\left(\frac{1.4\times.4}{2}\times 1\right)-\left(\frac{1.4\times.4\times-.6}{6}\times0\right)$$

$$= 55-3.2+.28-0 = 52.08 \text{ thousand rupees.}$$

Thus, the estimated value of sales for 1968 are 52.08 thousand rupees.

BINOMIAL EXPANSION METHOD

PROBLEM 8.30. *Estimate the outflow of silver from India in 1967-68 from the data given below.*

Average value of net exports of silver coins and bullion.

Year	*Rs.*
1963-64	*89,56,32,418*
1964-65	*52,53,74,607*
1965-66	*37,35,58,955*
1966-67	*27,84,61,129*
1968-69	*23,26,02,668*

SOLUTION. Estimation of the outflow of gold from India in 1967-68.

Years	Average value of net exports of gold coin and bullion in rupees	
1963-64	89,56,32, 418	y_0
1964-65	52,53, 74,607	y_1
1965-66	37, 35, 58,955	y_2
1966-67	27,84,61,129	y_3
1967-68	?	y_4
1968-69	23,26,02,668	y_5

Since the known figures are five, the fifth leading difference will be zero.

$\Delta^5_0 = y_5 - 5\,y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$

Substituting the above values, we get,

$\Delta^5_0 = 23{,}26{,}02{,}668 - 5y_4 + (10 \times 27{,}84{,}61{,}129)$

$-(10 \times 37{,}35{,}58{,}955) + (5 \times 52{,}53{,}74{,}607) - 89{,}56{,}32{,}418 = 0$

or $\Delta^5_0 = 23{,}26{,}02{,}668 - 5y_4 + 2{,}78{,}46{,}11{,}290$

$-3{,}73{,}55{,}89{,}550 + 2{,}62{,}68{,}73{,}035 - 89{,}56{,}32{,}418 = 0$

or $5y_4 = 23{,}26{,}02{,}668 + 2{,}78{,}46{,}11{,}290 - 3{,}73{,}55{,}89{,}550$

$+2{,}62{,}68{,}73{,}035 - 89{,}56{,}32{,}418$

or $5y_4 = 1{,}01{,}28{,}65{,}025$

or $y_4 = 20{,}25{,}73{,}005$.

Thus, the value of the estimated outflow of gold from India in 1967-68 is Rs. 20,25,73,005.

Note. Since the rate of decline in the value of silver exported is very high in the beginning than in the later years its cumulative effect has been reflected, in the interpolated figure. If the decline had been uniform, the interpolated figure for the year 1967-68 would not have been less than the figure of 1968-69.

PROBLEM 8.31. *Below are given weighted index numbers of cost of living of labourers in an industrial centre in India. Interpolate to find out the missing index number for 1973 to the nearest integer, using all the figures :*

Year	*Index*
1970	*173*
1971	*149*
1972	*145*
1974	*131*
1975	*141*

SOLUTION. Interpolation of the index number of cost of living for 1973.

Year	Index
1970	173 y_0
1971	149 y_1
1972	145 y_2
1973	... y_3
1974	131 y_4
1975	141 y_5

Since known figures are five, the fifth leading difference will be zero.

$\Delta^5_0 = y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0.$

Substituting the above values, we get

or $\Delta^5_0 = 141 - (5 \times 131) + 10y_3 - (10 \times 145) + (5 \times 149) - 173 = 0$

or $\Delta^5_0 = 141 - 655 + 10y_3 - 1450 + 745 - 173 = 0$

or $10y_3 = 655 - 141 + 1450 - 745 + 173 = 1392.$

or $y_3 = 139.2$

thus, the index number of cost of living for 1973 is 139.2

PROBLEM 8.32. *The age of mothers and the average number of children born per mother are given in a table below. Interpolate the average number of children born per mother aged 30-34.*

Age of mother in years	*No. of children born*
15—19	*0.7*
20 24	*2.1*
25—29	*3.5*
30—34	...
35—39	*5.7*
40—44	*5.8*

SOLUTION. Estimation of the average number of children born per mother aged 30-34.

Age of mother in years	*No. of children born*
15—19	0.7 y_0
20—24	2.1 y_1
25—29	3.5 y_2
30—34	... y_3
35—39	5.7 y_4
40—44	5.8 y_5

Since the known figures are five, the fifth leading difference will be zero.

$$\Delta^5{}_0 = y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$$

Substituting the given values, we get

$$\Delta^5{}_0 = 5.8 - (5 \times 5.7) + 10y_3 - (10 \times 3.5) + (5 \times 2.1) - 0.7 = 0$$

$$= 5.8 - 28.5 + 10y_3 - 35 + 10.5 - 0.7 = 0$$

or $10y_3 = 28.5 - 5.8 + 35 - 10.5 + 0.7 = 47.9$ or $y_3 = 4.79$

Thus, the expected average number of children born per mother aged 30—34 is 4.8 approx.

PROBLEM 8.33. *The following table gives the population of a town at the time of six censuses.*

1911	*75,401*
1921	*82,984*
1931	*86,686*
1941	*44,947*
1951	*93,091*
1961	*1,27,327*

Estimate the population for 1971.

SOLUTION. In the given data the population is continuously increasing from 1911 to 1931, but in 1941 it suddenly declines and again in 1951 it increases at once. So before an estimation of the population for 1971 is made, it is necessary to interpolate the population for 1941 as according to the assumptions underlying Newton's method of leading differences, there must be a gradual increase or decrease in the series.

Interpolation of the population for 1941.

Years	*Population*
1911	75,401 y_0
1921	82,984 y_1
1931	86,686 y_2
1941	? y_3
1951	93,091 y_4
1961	1,27,327 y_5

Since the known figures are five, the fifth leading difference will be zero.

$$\Delta^5{}_0 = y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$$

Substituting the above values we get,

$$\Delta^5{}_0 = 127327 - (5 \times 93091) + 10y_3 - (10 \times 86686) + (5 \times 82984) - (75401) = 0$$

$$= 127327 - 465455 + 10y_3 - 866860 + 414920 - 75401 = 0$$

or $10y_3 = 465455 + 866860 + 75401 - 127327 - 414920 = 865469$

or $y_3 = 86547$

∴ Population in 1941 = 86547, so the new table for interpolation of population in 1971 will be :

Years	*Population*
1911	75401 y_0
1921	82984 y_1
1931	86686 y_2
1941	86547 y_3
1951	93091 y_4
1961	127327 y_5
1971	…… y_6

Now the known quantities are six, therefore, the sixth leading difference will be zero.

$$\Delta^6{}_0 = y_6 - 6y_5 + 15y_4 - 20y_3 + 15y_2 - 6y_1 + y_0 = 0$$

Substituting the above values, we get

$$\Delta^6{}_0 = y_6 - (6\times127327) + (15\times93091) - (20\times86547)$$
$$+ (15\times86686) - (6\times82984) + 75401 = 0$$
$$y_6 = -763962 + 1396365 - 1730940 + 1300290 - 497904 + 75401 = 0.$$

or
$$y_6 = 763962 - 1396365 + 1730940 - 1300290 + 497904 - 75401$$
$$= 220750.$$

Note. Here again the sudden jump of population figure in 1961 has affected the extrapolated figure which is unusually high. The question could be solved by interpolating the figure of 1961 as well. The algebraic method of differences rests on the presumption that ultimately the differences would vanish and in cases where differences go on increasing, such inconsistent results are obtained.

PROBLEM 8.34. *Obtain the missing figure in the following table of one per cent values of chi-square and check the result by an algebraic method.*

Degree of Freedom	*1,*	*2,*	*3,*	*4,*	*5,*	*6,*	*7,*	*8,*	*9,*
1% chi-square	*6.64,*	*9.21,*	*11.34,*	*13.28*	*?,*	*16.81*	*18·48*	*20.07,*	*21.67*

SOLUTION. Estimation of the one per cent value of chi-square or five degrees of freedom.

Degrees of Freedom	*1% chi-square*
1	6.64 y_0
2	9.21 y_1
3	11.34 y_2
4	13.28 y_3
5	? y_4
6	16.81 y_5
7	18.48 y_6
8	20.07 y_7
9	21.67 y_8

Since the known quantities are eight the eighth leading difference will be zero.

$$\Delta^8_0 = y_8 - 8y_7 + 28y_6 - 56y_5 + 70y_4 - 56y_3 + 28y_2 - 8y_1 + y_0 = 0$$

Substituting the given values, we get

$$\Delta^8_0 = (21.67) - (8 \times 20.07) + (28 \times 18.48) - (56 \times 16{\cdot}81) + 70y_4$$
$$- (56 \times 13.28) + (28 \times 11.34) - (8 \times 9.21) + 6.64 = 0$$
$$= 21.67 - 160.56 + 517.44 - 941.36 + 70y_4 - 743.68$$
$$+ 317.52 - 73.68 + 6.64 = 0$$

or $70y_4 = 160.56 - 21.67 - 517.44 + 941.36 + 743.68 - 317.52 + 73.68 - 6.64$

$70y_4 = 1056.01$ or $y_4 = 15.09$.

Thus, the estimated value of 1% chi-square for five degrees of freedom is 15.09.

PROBLEM 8.35. *The following table gives premiums payable annually to secure an endowment policy of Rs. 2,000 with profits payable at the age of 60 or at death if it occurs earlier.*

Age at birthday	*Annual Premium* *Rs.*
20	*46.25*
21	*47.62*
22	*49.13*
23	*50.75*
24	*52.50*
25	*54.25*
26	*? ?*
27	*58.13*
28	*60.25*
29	*62.62*

SOLUTION. Estimation of the premium payable at the age of 26.

Age at birthday	*Annual Premium in rupees*
20	$46.25 \; y_0$
21	$47.62 \; y_1$
22	$49.13 \; y_2$
23	$50.75 \; y_3$
24	$52.50 \; y_4$
25	$54.25 \; y_5$
26	$? \; y_6$
27	$58.13 \; y_7$
28	$60.25 \; y_8$
29	$62.62 \; y_9$

Since known figures are nine, the ninth leading difference will be zero.

$$\Delta^9_0 = y_9 - 9y_8 + 36y_7 - 84y_6 + 126y_5 - 126y_4 + 84y_3 - 36y_2 + 9y_1 - y_0 \doteq 0$$

Substituting the given values, we have

$$\triangle^9{}_0 = 62.62-(9\times60.25)+(36\times58.13)-84y_6+(126\times54.25)-(126\times52.50)$$
$$+(84\times50.75)-(36\times49.13)+(9\times47.62)-46.25=0$$
$$=62.62-542.25+2092.68-84y_6+6835.50$$
$$-6615.00+4263.00-1768.68+428.58-46\ 25=0$$
$$84y_6=6262-542.25+2092.68+6835.50-6615.00+4263.00-1768.68+428.58$$
$$-46.25=4710.20$$
$$y_6=56.07\text{ Rs.}$$

Thus, the premium payable at the age of 26 for the said policy is Rs. 56.07

PROBLEM 8.36. *The number of students in a recognised institution of India for the last nine years is given below. Estimate the number for the year 1971-1972.*

Year	*No. of students*
1962-63	*1110*
1963-64	*1150*
1964-65	*1180*
1965-66	*1200*
1966-67	*1210*
1967-68	*1225*
1968-69	*1250*
1969-70	*1250*
1970-71	*1260*

SOLUTION. Estimation of the number of students for the year 1971-72.

Years	*No. of students*
1962-63	1110y_0
1963-64	1150y_1
1964-65	1180y_2
1965-66	1200y_3
1966-67	1210y_4
1967-68	1225y_5
1968-69	1250y_6
1969-70	1250y_7
1970-71	1260y_8
1971-72	? y_9

Since the known figures are nine, the ninth leading difference will be zero.

$$\triangle^9{}_0=y_9-9y_8+36y_7-84y_6+126y_5-126y_4+84y_3-36y_2+9y_1-y_0=0$$

Substituting the given values, we have

$$\triangle^9{}_0=y_9-(9\times1260)+(36\times1250)-(84\times1250)+(126\times1225)$$
$$-(126\times1210\ )+(84\times1200)-(36\times1180)+(9\times1150)-(1110)=0$$
$$=y_9-11340+45000-105000+154350-152460$$
$$+008000-42480+10350-1110=0$$

or
$$y_9=11340-45000+105000-154350+152460$$
$$-100800+42480-10350+1110=1890$$

Thus, the estimated number of students for the year 1971-72 is 1890.

Note. Here again the series is such that the differences would not vanish and so the extrapolated figure is unusually high. It would be better to solve this question by taking only the figures of 1969-70 and 1970-71, as in the year 1968-69 the figure is the same as in 1969-70.

PROBLEM 8.37 *It is required to find the missing value in the follo table. Establish any suitable formula for interpolation, and find the mis value.*

Serial No.	*Value*
1	*6.4577*
2	*3.4531*
3	*2.5604*
4	*2.1521*
5	*....*
6	*1.7849*
7	*1.6874*
8	*1.6177*
9	*1.5646*
10	*1.5232*

SOLUTION. Estimation of the value of Item No. 5.

Serial No.	*Value*
1	6.4577 y_0
2	3.4531 y_1
3	2.5604 y_2
4	2.1521 y_3
5	? y_4
6	1.7849 y_5
7	1.6874 y_6
8	1.6177 y_7
9	1.5646 y_8
10	1.5232 y_9

Since known quantities are nine, the ninth leading difference will be zero.

$$\Delta^9_0 = y_9 - 9y_8 + 36y_7 - 84y_6 + 126y_5 - 126y_4 + 84y_3 - 36y_2 + 9y_1 - y_0 = 0$$

Substituting the given values, we get,

$$\begin{aligned}\Delta^9_0 &= 1.5232 - (9 \times 1.5646) + (36 \times 1.6177) - (84 \times 1.6874) + (126 \times 1.7849) \\ &\quad - 126y_4 + (84 \times 2.1521) - (36 \times 2.5604) + (9 \times 3.4531) - 6.4577 = 0 \\ &= 1.5232 - 14.0814 + 58.2372 - 141.7416 + 224.8974 \\ &\quad - 126y_4 + 180.7764 - 92.1744 + 31.0779 - 6.4577 = 0\end{aligned}$$

or $$\begin{aligned}126y_4 &= 1.5232 - 14.0814 + 58.2372 - 141.7416 + 224.8974 + 180.7764 \\ &\quad - 92.1744 + 31.0779 - 6.4577 = 242.0579\end{aligned}$$

or $y_4 = 1.9211.$

Thus the interpolated value of the item No. 5 is 1.9211.

PROBLEM 8.38. *Interpolate the missing figures in the following table rice cultivation :—*

Year	*Acres in millions*
1961	*76.6*
1962	*78.7*
1963	*?*
1964	*77.7*
1965	*78.7*
1966	*?*
1967	*80.6*
1968	*77.6*
1969	*78.7*

SOLUTION. Estimation of the acres of land under rice cultivation for 63 and 1966.

Year	*Acres in millions*
1961	$76.6 y_0$
1962	$78.7 y_1$
1963	? y_2
1964	$77.7 y_3$
1965	$78.7 y_4$
1966	? y_5
1967	$80.6 y_6$
1968	$77.6 y_7$
1969	$78.7 y_8$

Known figures are seven. Therefore, the seventh first leading difference will e zero and similarly the seventh second leading difference will be zero.

$$\Delta^7{}_0 = y_7 - 7y_6 + 21y_5 + 35y_4 + 35y_3 - 21y_2 + 7y_1 - y_0 = 0 \ldots\ldots (i)$$

$$\Delta^7{}_1 = y_8 - 7y_7 + 21y_6 - 35y_5 + 35y_4 - 21y_3 + 7y_2 - y_1 = 0 \ldots\ldots (ii)$$

Multiplying (*ii*) by 3 we have

$$3y_8 - 21y_7 + 63y_6 - 105y_5 + 105y_4 - 63y_3 + 21y_2 - 3y_1 = 0. \ldots (iii)$$

Adding (*i*) & (*iii*) we have

$$3y_8 - 20y_7 + 56y_6 - 84y_5 + 70y_4 - 28y_3 + 4y_1 - y_0 = 0 \ldots\ldots\ldots (iv)$$

Substituting the table values in (*iv*) we have

$$(3\times78.7) - (20\times77.6) + (56\times80.6) - 84y_5 + (70\times78.7)$$
$$-(28\times77.7)\times(4\times78.7) - 76.6 = 0$$

or $84y_5 = 236.1 - 1552 + 4513.6 + 5509 - 2175.6 + 314.8 - 76.6 = 6769.3$

or $y_5 = 80.5$

Substituting the value of y in equation (*i*) we have

$$77{\cdot}6 - (7\times80.6) + (21\times80.5) - (35\times78.7) + (35\times77.7) - 21y_2 + (7\times78.7)$$
$$-76.6 = 0$$

or $77.6 - 564.2 + 1690.5 - 2754.5 + 2719.5 - 21y_2 + 550.9 - 76.6 = 0$

or $21y_2 = 1643.2$ or $y_2 = 78.2$

Thus, the estimated acres of land under rice cultivation in 1963 and 1966 re 78.2 millions and 80.5 millions respectively.

PROBLEM 8.39. *Interpolate the missing figure in the following table with he help of a suitable formula* :

Year	
1966	*1331*
1967	*1728*
1968	*2197*
1969	?
1970	*3375*
1971	*4096*
1972	*4913*

SOLUTION. Interpolation of the figure for the year 1969.

Year	
1966	$1331 y_0$
1967	$1728 y_1$
1968	$2197 y_2$
1969	$\ldots . y_3$
1970	$3375 y_4$
1971	$4096 y_5$
1972	$4613 y_6$

Since the known figures are six, the sixth leading difference would be zero

$$\triangle^6_0 = y_6 - 6y_5 + 15y_4 - 20y_3 + 15y_2 - 6y_1 + y_0 = 0$$

Substituting the above values, we get

$$\triangle^6_0 = 4913 - 24576 + 50625 - 20y_3 + 32955 - 10368 + 1331 = 0$$

$$20y_3 = 4913 - 24576 + 50625 + 32955 - 10368 + 1331$$

$$20y_3 = 54880$$

$$y_3 = 2744$$

Thus, the missing figure is 2744.

PROBLEM 8.40. *The following values are given in a table.*

x	*y*
1	*216,000*
2	*226,981*
3	*......*
4	*250,047*
5	*262,144*

Using any suitable algebraic method, find the value of y for x = 3.

Also draw a graph of the above points on a piece of squared paper and from this graph find the value of y for x = 4.4.

SOLUTION. Estimation of the value of y for $x=3$

x	y	
1	2,16,000	y_0
2	2,26,981	y_1
3	?	y_2
4	2,50,047	y_3
5	2,62,144	y_4

Since the known quantities are four, the fourth leading difference will be zero.

$$\triangle^4_0 = y_4 - 4y_3 + 6y_2 - 4y_1 + y_0 = 0$$

Substituting the given values, we get

$$\triangle^4_0 = 2{,}62{,}144 - (4 \times 2{,}50{,}047) + 6y_2 - (4 \times 2{,}26{,}981) + 2{,}16{,}000 = 0$$

$$= 2{,}62{,}144 - 10{,}00{,}188 + 6y_2 - 9{,}07{,}924 + 2{,}16{,}000 = 0$$

or $6y_2 = 2{,}62{,}144 - 10{,}00{,}188 - 9{,}07{,}924 + 2{,}16{,}000$

or $6y_2 = -2{,}62{,}144 + 10{,}00{,}188 + 9{,}07{,}924 - 2{,}16{,}000$

or $6y_2 = 19{,}08{,}112 - 4{,}78{,}144$

or $6y_2 = 14{,}29{,}968$

or $y_2 = 2{,}38{,}328$

Thus, the estimated value of y for $x=3$ is 2,38,328.

PROBLEM 8.41. *Find by algebraic method of interpolation, using all the information given, the likely number for 1970 from the following table of index numbers of production of a certain article in India :—*

Year	*Index number*
1968	*100*
1969	*107*
1970	..
1971	*157*
1972	*212*

SOLUTION. Estimation of the index number for the year 1970.

Year	*Index number*
1968	100 y_0
1969	107 y_1
1970	? y_2
1971	157 y_3
1972	212 y_4

Since the known figures are four, the fourth leading difference will be zero.

$$\triangle^4_0 = y_4 - 4y_3 + 6y_2 - 4y_1 + y_0 = 0$$

Substituting the above values, we get

$$\triangle^4_0 = 212 - (4 \times 157) + 6y_2 - (4 \times 107) + 100 = 0$$

$$= 212 - 628 + 6y_2 - 428 + 100 = 0$$

or $6y_2 = 212 - 628 - 428 + 100$

or $6y_2 = 628 + 428 - 212 - 100$

or $6y_2 = 744$

or $y_2 = 124$

Thus, the index number of production of a certain article for 1970 is 124.

PROBLEM 8.42. *The annual sales of a concern are given below :—*

Years	*Sales of cloth in million metres*
1950	*125*
1955	*163*
1960	*204*
1965	*238*
1970	*282*

Assuming the conditions of the market to be the same, estimate the sales for the year 1975.

SOLUTION. Extrapolation of the annual sale for the year 1975.

Years	*Sales (in lakhs of yards)*
1950	125 y_0
1955	163 y_1
1960	204 y_2
1965	238 y_3
1970	282 y_4
1975	.. y_5

Since the known figures are five the fifth leading difference would be zero.

$$\triangle^5_0 = 0$$

or $y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$

Substituting the given values, we get

$$\triangle^5_0 = y_5 - 1410 + 2380 - 2040 + 815 - 125 = 0$$

or $y_5 = 1410 - 2380 + 2040 - 815 + 125$

or $y_5 = 380$ million metres

Thus, the extrapolated sales for the year 1975 are 380 million metres.

LAGRANGE 'S METHOD

PROBLEM 8.43. *Determine by Lagrange's formula the national income of a country for the year 1972 from the following data :—*

National Income

Years	*Million Dollars*
1970	*5112*
1971	*6514*
1972	*?*
1973	*9069*
1974	*9685*

SOLUTION. Computation of the national income for 1972.

Years (x)	*Canadian National Income in millions of dollars* (y)
1970 x_0	5112 y_0
1971 x_1	6514 y_1
1973 x_2	9069 y_2
1974 x_3	9685 y_3

x=1972—The year for which the national income has to be interpolated. Applying to the above data, the Lagrange's formula of interpolation.

$$yx=y_0\frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}+y_1\frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}$$
$$+y_2\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}+y_3\frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}$$

where yx is the quantity to be interpolated, x is the given quantity in the variable corresponding to which yx is to be interpolated. $x_0; x_1, x_2, x_3 \ldots\ldots$are the given values of the variable x, and $y_0, y_1, y_2, y_3 \ldots\ldots$are the corresponding given values of the variable y.

We get,

$$yx=5112\frac{(1972-1971)(1972-1973)(1972-1974)}{(1970-1971)(1970-1973)(1970-1974)}$$
$$+6514\frac{(1972-1970)(1972-1973)(1972-1974)}{(1971-1970)(1971-1973)(1971-1974)}$$
$$+9069\frac{(1972-1970)(1972-1971)(1972-1974)}{(1973-1970)(1973-1971)(1973-1974)}$$
$$+9685\frac{(1972-1970)(1972-1971)(1972-1973)}{(1974-1970)(1974-1971)(1974-1973)}$$

or $$yx=5112\frac{(1)(-1)(-2)}{(-1)(-3)(-4)}+6514\frac{(2)(-1)(-2)}{(1)(-2)(-3)}$$
$$+9069\frac{(2)(1)(-2)}{(3)(2)(-1)}+9685\frac{(2)(1)(-1)}{(4)(3)(1)}$$

or $yx=-852+4342.67+6046-1614.16$

$=7922.51$ million dollars.

Thus, the estimated National Income for 1972 is 7922.51 million dollars.

PROBLEM 8.44. *The observed values of a function are respectively 168, 120, 72, and 63 at the four positions 3, 7, 9, and 10 of the independent variable. What is the estimate you can give for the value of the function at the position 6 of the independent variable ?*

SOLUTION. Estimation of the value of the function at the position 6 of the independent variable.

Independent variable (x)	*Values of the function* (y)
3 x_0	168 y_0
7 x_1	120 y_1
9 x_2	72 y_2
10 x_3	63 y_3

$x=6$, at which position of the independent variable, the value of the function has to be interpolated.

Applying Lagrange's formula, we get,

$$yx=168\,\frac{(6-7)\,(6-9)\,(6-10)}{(3-7)\,(3-9)\,(3-10)}+120\,\frac{(6-3)\,(6-9)\,(6-10)}{(7-3)\,(7-9)\,(7-10)}$$

$$+\,72\,\frac{(6-3)\,(6-7)\,(6-10)}{(9-3)\,(9-7)\,(9-10)}+63\,\frac{(6-3)\,(6-7)\,(6-9)}{(10-3)\,(10-7)\,(10-9)}$$

or $$yx=168\,\frac{(-1)\,(-3)\,(-4)}{(-4)\,(-6)\,(-7)}+120\,\frac{(3)\,(-3)\,(-4)}{(4)\,(-2)\,(-3)}+72\,\frac{(3)\,(-1)\,(-4)}{(6)\,(2)\,(-1)}$$

$$+\,63\,\frac{(3)\,(-1)\,(-3)}{(7)\,(3)\,(1)}$$

or $yx=12+180-72+27=147.$

Thus, the best estimate that can be given for the value of the function at the position 6 of the independent variable is 147.

PROBLEM 8.45. *In the following table h is the height above sea level and p the barometric pressure. Calculate p when h=5280.*

h = 0,	*4763,*	*6942,*	*10594*	
p =27,	*25,*	*23,*	*20*	

SOLUTION. Estimation of the barometric pressure when the height above sea level is 5280.

Height above sea level (x)	*Barometric pressure* (y)
0 x_0	27 y_0
4763 x_1	25 y_1
6942 x_2	23 y_2
10593 x_3	20 y_3

$x=5280$, the height above sea level for which the barometric pressure has to be interpolated.

Applying the Lagrange's Formula, we get,

$$yx=27\,\frac{(5280-4763)\,(5280-6942)\,(5280-10593)}{(0-4763)\,(0-6942)\,(0-10593)}$$

$$+\,5\,\frac{(5280-0)\,(5280-6942)\,(5280-10593)}{(4763-0)\,(4763-6942)\,(4763-10593)}$$

$$+23\,\frac{(5280-0)\,(5280-4763)\,(5280-10593)}{(6942-0)\,(6942-4763)\,(6942-10593)}$$

$$+\,20\,\frac{(5280-0)\,(5280-4763)\,(5280-6942)}{(10593-0)\,(10593-4763)\,(10593-6942)}$$

or $$yx=27\,\frac{(517)\,(-1662)\,(-5313)}{(-4763)\,(-6942)\,(-10593)}+25\,\frac{(5280)\,(-1662)\,(-5313)}{(4763)\,(-2179)\,(-5830)}$$

$$+23\,\frac{(5280)\,(517)(-5313)}{(6942)\,(2179)(-3651)}+20\,\frac{(5280)\,(517)\,(-1662)}{(10593)\,(5830)\,(3651)}$$

or $yx=-.353+19.23+6.34-.402.$

or $yx=24.8$

Thus, the estimated barometric pressure when the height above sea level is 5280 is 24.8.

PROBLEM 8.46 *Given* $\log_{10} 654=2.8156$; $\log_{10} 658=2.8182$;
$\log_{10} 659=2.8189$; $\log_{10} 661=2.8202$;
Find $\log_{10} 656$.

SOLUTION. Computation of the value of $\log_{10} 656$.

(x)	(y)
$\log_{10} 654 x_0$	$2.8156\ y_0$
$\log_{10} 658\ x_1$	$2.8182\ y_1$
$\log_{10} 659\ x_2$	$2.8189\ y_2$
$\log_{10} 661\ x_3$	$2.8202\ y_3$

$x=\log_{10} 656$, *i. e.* the number for which the value has to be intered.

Applying to the above data, the Lagrange's formula of interpolation.

$$yx=y_0\,\frac{(x-x_1)\,(x-x_2)\,(x-x_3)}{(x_0-x_1)\,(x_0-x_2)\,(x_0-x_3)}+y_1\,\frac{(x-x_0)\,(x-x_2)\,(x-x_3)}{(x_1-x_0)\,(x_1-x_2)\,(x_1-x_3)}$$

$$+y_2\,\frac{(x-x_0)\,(x-x_1)\,(x-x_3)}{(x_2-x_0)\,(x_2-x_1)\,(x_2-x_3)}+y_3\,\frac{(x-x_0)\,(x-x_1)\,(x-x_2)}{(x_3-x_0)\,(x_3-x_1)\,(x_3-x_2)}$$

where yx is the quantity to be interpolated, x is the given quantity in the x variable corresponding to which yx is to be interpolated $x_0, x_1, x_2, x_3 \ldots\ldots$ the given values of the variable x, and $y_0, y_1, y_2, y_3, \ldots\ldots$ are the corresponding given values of the variable y.

We get,

$$yx=2.8156\,\frac{(656-658)\,(656-659)\,(656-661)}{(654-658)\,(654-659)\,(654-661)}$$

$$+2.8182\,\frac{(656-654)\,(656-659)\,(656-661)}{(658-654)\,(658-659)\,(658-661)}$$

$$+2.8189\,\frac{(656-654)\,(656-658)\,(656-661)}{(659-654)\,(659-658)\,(659-661)}$$

$$+2.8202\,\frac{(656-654)\,(656-658)\,(656-659)}{(661-654)\,(661-658)\,(661-659)}$$

$$yx=2.8156\,\frac{(-2)\,(-3)\,(-5)}{(-4)\,(-5)\,(-7)}+2.8182\,\frac{(2)\,(-3)\,(-5)}{(4)\,(-1)\,(-3)}$$

$$+2.8189\,\frac{(2)\,(-2)\,(-5)}{(5)\,(1)\,(-2)}+2.8202\,\frac{(2)\,(-2)\,(-3)}{(7)\,(3)\,(2)}$$

or $yx=.6033+7.0455-5.6378+.8059.$

or $yx=2.8169.$

Thus, the estimated value of $\log_{10} 656=2.8169$.

PROBLEM 8.47. *What do you understand by interpolation and extrapolation? What are their uses? The following table gives the normal weights of babies of a specie during the first twelve months of life :*

Age in months	*0*	*2*	*5*	*8*	*10*	*12*
Weights in kgs.	*$7\frac{1}{2}$*	*$10\frac{1}{4}$*	*15*	*16*	*18*	*21*

Estimate the weight of the baby at the age of 7 months.

SOLUTION. Estimation of the weight of the baby at the age of 7 months.

Age in months (x)		*Weight in pounds* (y)	
0	x_0	7.50	y_0
2	x_1	10.25	y_1
5	x_2	15.00	y_2
8	x_3	16.00	y_3
10	x_4	18.00	y_4
12	x_5	21.00	y_5

$x=7$, at which age the weight of the baby has to be found out. Substituting the above values in the Lagrange's formula :

$$yx = y_0\frac{(x-x_1)(x-x_2)(x-x_3)(x-x_4)(x-x_5)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)(x_0-x_4)(x_0-x_5)}$$
$$+y_1\frac{(x-x_0)(x-x_2)(x-x_3)(x-x_4)(x-x_5)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)(x_1-x_4)(x-x_5)}$$
$$+y_2\frac{(x-x_0)(x-x_1)(x-x_3)(x-x_4)(x-x_5)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)(x_2-x_4)(x_2-x_5)}$$
$$+y_3\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_4)(x-x_5)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)(x_3-x_4)(x_3-x_5)}$$
$$+y_4\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)(x-x_5)}{(x_4-x_0)(x_4-x_1)(x_4-x_2)(x_4-x_3)(x_4-x_5)}$$
$$+y_5\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)(x-x_4)}{(x_5-x_0)(x_5-x_1)(x_5-x_2)(x_5-x_3)(x_5-x_4)}$$

where yx is the quantity to be interpolated.

We get,

$$yx = 7.5\frac{(7-2)(7-5)(7-8)(7-10)(7-12)}{(0-2)(0-5)(0-8)(0-10)(0-12)}$$
$$+10.25\frac{(7-0)(7-5)(7-8)(7-10)(7-12)}{(2-0)(2-5)(2-8)(2-10)(2-12)}$$
$$+15\frac{(7-0)(7-2)(7-8)(7-10)(7-12)}{(5-0)(5-2)(5-8)(5-10)(5-12)}$$
$$+16\frac{(7-0)(7-2)(7-5)(7-10)(7-12)}{(8-0)(8-2)(8-5)(8-10)(8-12)}$$
$$+18\frac{(7-0)(7-2)(7-5)(7-8)(7-12)}{(10-0)(10-2)(10-5)(10-8)(10-12)}$$
$$+21\frac{(7-0)(7-2)(7-5)(7-8)(7-10)}{(12-0)(12-2)(12-5)(12-8)(12-10)}$$

or $$yx = 7.5\frac{(5)(2)(-1)(-3)(-5)}{(-2)(-5)(-8)(-10)(-12)} + 10.25\frac{(7)(2)(-1)(-3)(-5)}{(2)(-3)(-6)(-8)(-10)}$$
$$+15\frac{(7)(5)(-1)(-3)(-5)}{(5)(3)(-3)(-5)(-7)} + 16\frac{(7)(5)(2)(-3)(-5)}{(8)(6)(3)(-2)(-4)}$$

$$+18\ \frac{(7)(5)(2)(-1)(-5)}{(10)(8)(5)(2)(-2)}+21\ \frac{(7)(5)(2)(-1)(-3)}{(12)(10)(7)(4)(2)}$$

$$\text{or } yx = .12-.75+5+14.58-3.94+.65.$$

$$= 15.66.$$

Thus the weight of the baby at the age of 7 months is 15.66 kgs. appro Since the increase in weight in earlier months is disproportionately higher th in the latter months, the interpolated figure is rather incorrect.

PROBLEM 8.48. *Estimate the probable number of persons earning betw Rs. 30 and 40 from the following :—*

Income in Rs.	*Number of persons*
Rs. 15-20	*73*
Rs. 20-30	*97*
Rs. 30-45	*110*
Rs. 45-55	*180*
Rs. 55-70	*140*

SOLUTION. Estimation of the probable number of persons earning bel Rs. 40.

Income below rupees x	*Number of persons (cumulative frequency)* y
20 x_0	73 y_0
30 x_1	170 y_1
45 x_2	280 y_2
55 x_3	460 y_3
70 x_4	600 y_4

$x=40$, *i.e.* corresponding to which income the number of persons has to b interpolated.

Applying the Lagrange's Formula, we get

$$yx=73\ \frac{(40-30)(40-45)(40-55)(40-70)}{(20-30)(20-45)(20-55)(20-70)}$$

$$+170\frac{(40-20)(40-45)(40-55)(40-70)}{(30-20)(30-45)(30-55)(30-70)}$$

$$+280\frac{(40-20)(40-30)(40-55)(40-70)}{(45-20)(45-30)(45-55)(45-70)}$$

$$+460\frac{(40-20)(40-30)(40-45)(40-70)}{(55-20)(55-30)(55-45)(55-70)}$$

$$+600\frac{(40-20)(40-30)(40-45)(40-55)}{(70-20)(70-30)(70-45)(70-55)}$$

$$\text{or } yx=73\Big(\frac{(10)(-5)(-15)(-30)}{(-10)(-25)(-35)(-50)}+170\frac{(20)(-5)(-15)(-30)}{(10)(-15)(-25)(-40)}$$

$$+\ 280\ \frac{(20)(10)(-15)(-30)}{((25)(15)(-10)(-25)}+460\frac{(20)(10)(-5)(-30)}{(35)(25)\quad(10)(-15)}\Big)$$

$$+\ 600\ \frac{(20)(10)(-5)(-15)}{(50)(40)(25)(15)}$$

$$\text{or } yx= \quad -3.75+51+268.8-105.14+12 = 202.91$$

$$= \quad 223 \text{ persons.}$$

Thus, the number of persons earning below Rs. 40 is 223 and the number o persons earning below Rs. 30 is 170 (given).

Therefore the probable number of persons earning between Rs. 30 and 40 i 223—170=53.

PROBLEM 8.49. *Determine by Lagrange's formula the percentage number of criminals under 35 years.*

Age	*% number of criminals*
Under 25 years	*52.0*
,, 30 ,,	*67.3*
,, 40 ,,	*84.1*
,, 50 ,,	*94.4*

SOLUTION. Estimation of the percentage number of criminals under 35 years.

Age (x)	*% number of criminals* (y)
Under 25 years x_0	52.0 y_0
,, 30 ,, x_1	67.3 y_1
,, 40 ,, x_2	84.1 y_2
,, 50 ,, x_3	94.4 y_3

$x=35$, corresponding to which year the percentage number of criminals has to be interpolated.

Applying the Lagrange's formula, we get,

$$yx = 52.0\,\frac{(35-30)\,(35-40)\,(35-50)}{(25-30)\,(25-40)\,(25-50)}$$

$$+\,67.3\,\frac{(35-25)\,(35-40)\,(35-50)}{(30-25)\,(30-40)\,(30-50)}$$

$$+\,84.1\,\frac{(35-25)\,(35-30)\,(35-50)}{(40-25)\,(40-30)\,(40-50)}$$

$$+\,94.4\,\frac{(35-25)\,(35-30)\,(35-40)}{(50-25)\,(50-30)\,(50-40)}$$

$$\text{or } yx=52\,\frac{(5)\,(-5)\,(-15)}{(-5)\,(-15)\,(-25)}+67.3\,\frac{(10)\,(-5)\,(-15)}{(5)\,(-10)\,(-20)}$$

$$+84.1\,\frac{(10)\,(5)\,(-15)}{(15)\,(10)\,(-10)}\;+94.4\,\frac{(10)\,(5)\,(-5)}{(25)\,(20)\,(10)}$$

or $yx=-10.4+50.5+42.05-4.72$

or $yx=\;77.43.$

Thus, the estimated percentage number of criminals under 35 years is 77.43.

PROBLEM 8.50. *From the following figures estimate the number of persons whose income is Rs. 19 or more but does not exceed Rs. 25.*

Income in Rs.	*Number of persons*
1 and not exceeding 9	*50*
10 ,, ,, ,, 19	*70*
19 ,, ,, ,, 28	*203*
28 ,, ,, ,, 37	*406*
37 ,, ,, ,, 46	*304*
46 ,, ,, ,, 55	*42*
55 ,, ,, ,, 64	*5*

SOLUTION. Estimation of the number of persons whose income does not exceed Rs. 25.

Income not exceeding Rs. (x)	*Number of persons* (y)
9 x_0	50 y_0
19 x_1	120 y_1
28 x_2	323 y_2
37 x_3	729 y_3
46 x_4	1033 y_4
55 x_5	1075 y_5
64 x_6	1080 y_6

$x=25$, corresponding to which, the value of y has to be interpolated. Applying Lagrange's Formula, we get

$$\begin{aligned} x = {} & 50\,\frac{(25-19)\,(25-28)\,(25-37)\,(25-46)\,(25-55)\,(25-64)}{(9-19)\,(9-28)\,(9-37)\,(9-46)\,(9-55)\,(9-64)} \\ & + 120\,\frac{(25-9)\,(25-28)\,(25-37)\,(25-46)\,(25-55)\,(25-64)}{(19-9)\,(19-28)\,(19-37)\,(19-46)\,(19-55)\,(19-64)} \\ & + 323\,\frac{(25-9)\,(25-19)\,(25-37)\,(25-46)\,(25-55)\,(25-64)}{(28-9)\,(28-19)\,(28-37)\,(28-46)\,(28-55)\,(28-64)} \\ & + 729\,\frac{(25-9)\,(25-19)\,(25-28)\,(25-46)\,(25-55)\,(25-64)}{(37-9)\,(37-19)\,(37-28)\,(37-46)\,(37-55)\,(37-64)} \\ & + 1033\,\frac{(25-9)\,(25-19)\,(25-28)\,(25-37)\,(25-55)\,(25-64)}{(46-9)\,(46-19)\,(46-28)\,(46-37)\,(46-55)\,(46-64)} \\ & + 1075\,\frac{(25-9)\,(25-19)\,(25-28)\,(25-37)\,(25-46)\,(25-64)}{(55-9)\,(55-19)\,(55-28)\,(55-37)\,(55-46)\,(55-64)} \\ & + 1080\,\frac{(25-9)\,(25-19)\,(25-28)\,(25-37)\,(25-46)\,(25-55)}{(64-9)\,(64-19)\,(64-28)\,(64-37)\,(64-46)\,(64-55)} \end{aligned}$$

$$\begin{aligned} \text{Or } yx = {} & 50\,\frac{(6)\,(-3)\,(-12)\,(-21)\,(-30)\,(-39)}{(-10)\,(-19)\,(-28)\,(-37)\,(-46)\,(-55)} \\ & + 120\,\frac{(16)\,(-3)\,(-12)\,(-21)\,(-30)\,(-39)}{(10)\,(-9)\,(-18)\,(-27)\,(-36)\,(-45)} \\ & + 323\,\frac{(16)\,(6)\,(-12)\,(-21)(-30)\,(-39)}{(19)\,(9)\,(-9)\,(-18)\,(--27)\,(-36)} \\ & + 729\,\frac{(16)\,(6)\,(-3)\,(-21)\,(-30)\,(-39)}{(28)\,(18)\,(9)\,(-9)\,(-18)\,(-27)} \\ & + 1033\,\frac{(16)\,(6)\,(-3)\,(-12)\,(-30)\,(-39)}{(37)\,(27)\,(18)\,(9)\,(-9)\,(-18)} \\ & + 1075\,\frac{(16)\,(6)\,(-3)\,(-12)\,(-21)\,(-39)}{(46)\,(36)\,(27)\,(18)\,(9)\,(-9)} \\ & + 1080\,\frac{(16)\,(6)\,(-3)\,(-12)\,(-21)\,(-30)}{(55)\,(45)\,(36)\,(27)\,(18)\,(9)} \end{aligned}$$

Or $yx = -.18 + 23.98 + 332.31 - 260 + 150.93 - 46.7 + 6.03$

Or $yx = 206$.

Thus, the number of persons whose income does not exceed Rs. 25 is 206. and the number of persons whose income does not exceed Rs. 19 is 120.

Therefore, the number of persons whose income is Rs. 19 or more but does not exceed Rs. 25 is 206—120=86.

PROBLEM. 8.51. *The following are the marks obtained by 65 students in statistics in 1975 M. Com. (final examination) of a certain university :—*

Marks (out of 100)	*Number of students*
More than 25	*65*
,, ,, 36	*63*
,, ,, 45	*40*
,, ,, 55	*18*
,, ,, 70	*7*

Find out the number of students who secured 1st class marks in statistics.

SOLUTION. First class marks means 60 per cent or more. So we have to interpolate the number of students who secured 60 marks or more than 60 marks. Assuming that marks are not given in fractions, we may interpolate from the given data, the number of students who secured more than 59 marks.

Estimation of the number of students who secured more than 59 marks.

Marks more than *(out of 100)* (x)		*Number of students* (y)	
25	x_0	65	y_0
36	x_1	63	y_1
45	x_2	40	y_2
55	x_3	18	y_3
70	x_4	7	y_4

$x=59$, corresponding to which the value of y has to be interpolated.

Applying Lagrange's Formula, we get,

$$yx=65\,\frac{(59-36)\,(59-45)\,(59-55)\,(59-70)}{(25-36)\,(25-45)\,(25-55)\,(25-70)}$$

$$+63\,\frac{(59-25)\,(59-45)\,(59-55)\,(59-70)}{(36-25)\,(36-45)\,(36-55)\,(36-70)}$$

$$+40\,\frac{(59-25)\,(59-36)\,(59-55)\,(59-70)}{(45-25)\,(45-36)\,(45-55)\,(45-70)}$$

$$+18\,\frac{(59-25)\,(59-36)\,(59-45)\,(59-70)}{(55-25)\,(55-36)\,(55-45)\,(55-70)}$$

$$+\,7\,\frac{(59-25)\,(59-36)\,(59-45)\,(59-70)}{(70-25)\,(70-36)\,(70-45)\,(70-55)}$$

$$\text{or } yx=65\,\frac{(23)\,(14)\,(4)\,(-11)}{(-11)\,(-20)\,(-30)\,(-45)}+63\,\frac{(34)\,(14)\,(4)\,(-11)}{(11)\,(-9)\,(-19)\,(-34)}$$

$$+40\,\frac{(34)\,(23)\,(4)\,(-11)}{(20)\,(9)\,(-10)\,(-25)}+18\,\frac{(34)\,(23)\,(14)\,(-11)}{(30)\,(19)\,(10)\,(-15)}$$

$$+\,7\,\frac{(34)\,(23)\,(14)\,(4)}{(45)\,(34)\,(25)\,(15)}$$

$$\text{or } yx=-3.1+20.5-26.7+25.4+.54$$

$$=-16.64 \text{ or } 17.$$

Thus the estimated number of students who secured 1st class marks in statistics is 17.

DIVIDED DIFFERENCE METHOD

This method is used in case of unequal intervals. If y_0, y_1, y_2 etc. are the values of y corresponding to x_0, x_1, x_2 etc. a divided difference table will be constructed as follows :—

DIVIDED DIFFERENCE TABLE

x	y	1st	2nd	3rd
x_0	y_0			
		$\Delta^{1}{}_{1}=\frac{y_1-y_0}{x_1-x_0}$		
x_1	y_1		$\Delta^{2}{}_{1}=\frac{\Delta^{1}{}_{2}-\Delta^{1}{}_{1}}{x_2-x_0}$	
		$\Delta^{1}{}_{2}=\frac{y_2-y_1}{x_2-x_1}$		$\Delta^{3}{}_{1}=\frac{\Delta^{2}{}_{2}-\Delta^{2}{}_{1}}{x_3-x_0}$
x_2	y_2		$\Delta^{2}{}_{2}=\frac{\Delta^{1}{}_{3}-\Delta^{1}{}_{2}}{x_3=x_1}$	
		$\Delta^{1}{}_{3}=\frac{y_3-y_2}{x_3-x_2}$		
x_3	y_3			

PROBLEM 8.52. *Construct a divided difference table for the followin data :—*

x :	*2*	*4*	*5*	*8*
y :	*0.3010*	*0.6021*	*0.6990*	*0.9031*

SOLUTION

DIVIDED DIFFERENCE TABLE

		Divided Differences		
x	y	First Δ^1	Second Δ^2	Third Δ^3
$2x_0$	$0.3010y_0$			
		$\frac{0.6021-0.3010}{4-2}=\frac{0.3011}{2}$		
$4x_1$	$0.6021y_1$	$\Delta^1_1=0.15055$	$\frac{.09690-.15055}{5-2}=\frac{-.05365}{3}$	
		$\frac{0.6990-0.6021}{5-4}=\frac{0.0969}{1}$	$\Delta^2_1=-0.01788$	$\frac{-.00722+.01788}{8-2}$
$5x_2$	$0.6990y_2$	$\Delta^1_2=0.09690$	$\frac{.06803-.09690}{8-4}=\frac{-.02887}{4}$	$\Delta^3_1=\frac{.01066}{6}=0.001777$
		$\frac{0.9031-0.6990}{8-5}=\frac{0.2041}{3}$	$\Delta^2_2=-.007217$	
$8x_3$	$0.9031y_3$	$\Delta^1_3=0.06803$		

PROBLEM 8.53. *The following table lists the value of sinh x for certain values of x. Calculate from it approximate value of sinh x for x=0.23*

x	*0.00*	*0.20*	*0.30*	*0.50*
sinh x	*0.00000*	*0.20134*	*0.30452*	*0.52110*

SOLUTION .

		Divided Differences		
x	y	1st Δ^1	2nd Δ^2	3rd Δ^3
$x_0=0.00$	$0.00000y_0$	$\frac{0.20134}{0.20}=1.0067$	$\frac{0.0251}{0.30}=0.08367$	
$x_1=0.20$	$0.20134y_1$	$\frac{0.10318}{0.10}=1.0318$		$\frac{0.08666}{0.50}$
$x_2=0.30$	$0.30452y_2$	$\frac{0.21658}{0.20}=1.0829$	$\frac{0.0511}{0.30}=0.17033$	$=0.17332$
$x_3=0.50$	$0.52110y_3$			

The formula of divided differences as given by Newton is

$$r\ x=yo+(x-x_0)\Delta^1_1+(x-x_0)(x-x_1)\Delta^2_1+(x-x_0)(x-x_1)(x-x_2)\Delta^3_1$$

$$=0.00000+(.23—00)\,(1.0067)+(.23—.00)\,(.23—.20)\,(2.08367)$$
$$+(.23—.00)\,(.23—.20)\,(.23—.30)\,(.17332)$$
$$=.231541+.000577323—.000084$$
$$=.232034323.$$

PROBLEM 8.54.

Given $\log_{10} 654=2.8156$; $\log_{10} 658=2.8182$; $\log_{10} 659=2.8189$
$\log_{10} 661=2.8202$

Find $\log_{10} 656$ *using two different interpolations formulae available for observations at unequal intervals, say, Lagrange's formula and the formula for divided differences.*

SOLUTION. By divided differences :

		Divided Differences		
x	y	Δ^1	Δ^2	Δ^3
654 x_0	2.8156 y_0			
		$\Delta^1{}_1=\frac{2.8182—2.8156}{658—654}$ $=.00065$		
658 x_1	2.8182 y_1		$\Delta^2{}_1=\frac{.0007—.00065}{659—654}$ $=.00001$	
		$\Delta^1{}_3=\frac{2.8189—2.8182}{659—658}$ $=.0007$		$\Delta^3{}_1=\frac{—.000027}{7}$ $=.0000038$
659 x_2	2.8189 y_2		$\Delta^2{}_2=\frac{.00065—.0007}{661—658}$ $=—.000017$	
		$\Delta^1{}_3=\frac{2.8202—2.8189}{661—659}$ $=.00065$		
661 x_3	2.8202 y_3			

$$y_x=y_0+(x—x_0)^1\Delta_1+(x—x_0)\,(x—x_1)\Delta^2{}_1+(x—x_0)\,(x—x_1)\,(x—x_2)\Delta^3{}_1$$
$$y_{656}=2.8156+(2\times.00065)+(2\times—2\times.00001)+(2\times—2\times—3\times.0000038)=2.8169$$

The Lagrange's formula will also give the same answer.

PROBLEM 8.55. *The following table gives the values of cosh x for certain values of x. Calculate from it approximate values of cosh x for x=0.14.*

x	*0.10*	*0.12*	*0.15*	*0.16*	*0.18*
cosh x	*1.0050*	*1.0072*	*1.0113*	*1.0128*	*1.0162*

SOLUTION.

Difference Table

x	cosh x	Differences			
		First	Second	Third	Fourth
$x_0=0.10$	1.0050 y_0				
		$\frac{0.0022}{0.02}=0.11000$			
$x_1=0.12$	1.0072 y_1		$\frac{0.02333}{0.05}=0.46660$		
		$\frac{0.0041}{0.03}=0.13333$		$\frac{-0.04985}{0.06}=-0.83083$	
$x_2=0.15$	1.0113 y_2		$\frac{0.01667}{0.04}=0.41675$		$\frac{-3.33450}{.08}=-41.681$
		$\frac{0.0015}{0.01}=0.15000$		$\frac{-.24992}{0.06}=-4.16533$	
$x_3=0.16$	1.0128 y_3		$\frac{0.020000}{.03}=0.66667$		
		$\frac{0.0034}{0.02}=0.17000$			
$x=0.16$	1.0162 y_4				

$$
\begin{aligned}
yx &= y_0+(x-x_0)\Delta^1{}_1+(x-x_0)(x-x_1)\Delta^2{}_1+(x-x_0)(x-x_1)(x-x_2)\Delta^3{}_1 \\
&\qquad +(x-x_0)(x-x_1)(x-x_2)(x-x_3)\Delta^4{}_1 \\
&=1.0050+(.04\times.1100)+(.04\times.02\times.46660)+(.04\times.02\times-.01\times.83833) \\
&\qquad +(.04\times.02\times-.01\times-.02\times-41.5875) \\
&=1.0050+.004400+.000373-.0000067-.000000=1.0097.
\end{aligned}
$$

PROBLEM 8.56. *From the data given below interpolate the value of*
when $x=4$

x	*1*	*3*	*5*	*6*
y	*43833*	*46568*	*49431*	*50912*

SOLUTION

		Differences		
x	y	Δ^1	Δ^2	Δ^3
1 x_0	43833 y_0			
		1367.5 $\Delta^1{}_1$		
3 x_1	46568 y_1		16 $\Delta^2{}_1$	.1 Δ^3
		1431.5 y $\Delta^1{}_2$		
5 x_2	49431 y_2		16.5 $\Delta^2{}_2$	
		1481.0 $\Delta^1{}_3$		
6 x_3	50912 y_3			

$$y_x = y_0 + (x-x_0)\triangle^1{}_1 + (x-x_0)(x-x_1)\triangle^2{}_1 + (x-x_0)(x-x_1)(x-x_2)\triangle^3{}_1$$

Substituting the value we get

$$y_x = 43833 + (4-1)(1367.5) + (4-1)(4-3)(16) + (4-1)(4-3)(4-5)(.1)$$

$$= 43833 + 4102.5 + 48 - .3$$

$$= 47983.2.$$

REVISIONAL PROBLEMS – 8

8.57 Use any algebraic method to interpolate the population for 1951

Years	*Population (in lakhs)*
1931	2,528
1941	2,514
1961	3,168
1971	3,613

(*2765.7 lakhs*)

8.58 Use one of the methods of interpolation to estimate the business done in 1970 from the following data :

Years	*Business done (Rs. in lakhs)*
1967	150
1968	235
1969	365
1971	525
1972	780

(*Rs. 447 lakhs*)

8.59. Interpolate the missing figures in the following table with the help of a suitable formula :

Years	Acres *in millions*
1961	1331
1962	1728
1963	2197
1964	?
1965	3375
1966	4096
1967	4913

(*2744 Million Acres*)

8.60. Use some appropriate interpolation method and reconstruct the following frequency table with the intervals halved :

x	*Frequency*
0-2	35
2-4	52
4-6	84

[	*0-1,*	*1-2,*	*2-3,*	*3-4,*	*4-5,*	*5-6*
F	*21*	*14*	*22*	*30*	*38*	*46*]

8.61. The observed values of a function are respectively 168, 120, 72 and 63 at the four functions 3, 7, 9 and 10 of the independent variable. What is the best estimate you can give for the value of the function at the position 6 of the independent variable ? Use Lagrange's formula. (*147*)

8.62. The following are the amounts of income tax paid by a few businessmen during one year :

More than	Rs. 500	600
,,	,, 1000	550
,,	,, 1500	425
,,	,, 2000	275
,,	,, 2500	100
,,	,, 3000	25

Find out the number of businessmen who paid more than Rs. 1200 but more than Rs. 2400 as income tax. (*369 person*

8.63. From the following data estimate the number of persons in the inco group Rs. 20 and Rs. 25 :

Income	*No. of persons.*
Below Rs. 10	20
,, Rs. 20	45
,, Rs. 30	115
,, Rs. 40	210
,, Rs. 50	325

(3

8.64. From the following data of the population of a city in lakhs find out population for 1956 by the Parabolic curve method :

Year	1941	1951	1961	1971
Population in lakhs	18	22	25	30

(23.4

8.65. Use Newton's formula and interpolate the migration from Town A in year 1963 from the following :

Years	1962	1964	1966	1968	1970
Migration from Town A	233	391	582	799	1035

(*307.44* or 307

Association of Attributes 9

Association of attributes refers to a study of the relationship between two attributes, having class-frequencies. The method of correlation is employed to study the relationship between two series which are capable of quantitative measurement and which can be expressed in class-intervals and frequencies. There are, however, many things which are incapable of being measured in figures, *e.g.*, blindness, and in which, only the presence or absence of a phenomenon can be studied. It is in such cases that the method of Association of Attributes is applied to study the relationship between two phenomena.

In this chapter questions have been solved first on the classification of data and then on the study of association. Questions of association have been solved in any of the three ways, *viz.*, by comparing the actual and the expected frequencies, by comparing the percentages or proportions of the association of say attribute A in the universe of B and β, and by calculating the Co-efficient of Association.

CLASSIFICATION OF DATA

PROBLEM 9.1. *Given the following frequencies of the positive classes, find the frequencies of the ultimate classes.*

$(A)=80,\ (B)=100,\ (AB)=70,\ N=250$

SOLUTION. The ultimate classes are
$(AB),\ (A\beta),\ (\alpha B),\ (\alpha\beta)$; $(AB)=70$ (given) ; $(A\beta)=(A)-(AB)$;

$(\alpha B)=(B)-(AB)$
$(\alpha\beta)=(\alpha)-(\alpha B)=N-(A)-(B)+(AB)$
Substituting the given values, we get
$(AB)=70$; $(A\beta)=80-70=10.$
$(\alpha B)=100-70=30$; $(\alpha\beta)=250-80-100+70=140.$

Thus the ultimate class frequencies are $(AB)=70,\ (A\beta)=100\ (\alpha B)=30,$ $(\alpha\beta)=140.$

PROBLEM 9.2. *Given the following ultimate class-frequencies, and the frequencies of the positive and negative classes and the whole number of observations N.*

$(AB)=250$; $(A\beta)=120$; $(\alpha B)=200$; $(\alpha\beta)=70.$

SOLUTION. The classes for which frequencies have to be found are :

$N,\ (A),\ (B),\ (\alpha),\ (\beta)$

Now,

$N=(AB)+(A\beta)+(\alpha B)+(\alpha\beta)$;

$(A)=(AB)+(A\beta)$; $(B)=(AB)+(\alpha B)$
$(\alpha)=(\alpha B)+(\alpha\beta)$; $(\beta)=(A\beta)+(\alpha\beta)$

Substituting the given values, we get,

$N=250+120+200+70=640$
$(A)=250+120=370$; $(B)=250+200=450$
$(\alpha)=200+70=270$; $(\beta)=120+70=190$
Thus, $N=640,\ (\alpha)=270$: $(A)=370,\ (\beta)=190$, and, $(B)=450$

PROBLEM 9.3. *Given the following frequencies of the positive classes find all the class-frequencies.*

$N=12{,}000\ (AB)=453\ ;\ (A)=977;(AC)=284\ ;$

$(B)=1185;(BC)=250\ ;\ (C)=596;(ABC)=127$

SOLUTION. The classes for which frequencies have to be found out are :

$(a), (\beta), (\gamma), (A\beta), (aB), (A), (aC), (a\gamma),$
$(B\gamma), (\beta C), (\beta\gamma), (AB\gamma), (A\beta C), (aBC), (A\beta\gamma), (aB\gamma),$
$(a\beta C), (a\beta\gamma).$

Now,

$(a)=N-(A)\ ;\ (\beta)=N-(B)$
$(\gamma)=N-(C)\ ;\ (AB\gamma)=(AB)-(ABC)$
$(A\beta C)=(AC)-(ABC)\ ;\ (aBC)=(BC)-(ABC)$
$(A\beta\gamma)=(A\beta)-(A\beta C)=(A)-(AB)-(A\beta C)$
$(aB\gamma)=(aB)-(aBC)=(B)-(AB)-(aBC)$
$(a\beta C)=(\beta C)-(A\beta C)=(C)-(BC)-(A\beta C)$
$(A\beta)=(A\beta C)+(A\beta\gamma)\ ;\ (aB)=(aBC)+(aB\gamma)$
$(A\gamma)=(AB\gamma)+(A\beta\gamma)\ ;\ (aC)=(aBC)+(a\beta C)\ ;\ (B\gamma)=(AB\gamma)+(aB\gamma)$
$(\beta C)=(A\beta C)+(a\beta C)\ ;\ (a\beta\gamma)=(a\beta)-(a\beta C)=(\beta)-(A\beta)-(a\beta C)$
$(a\beta)=(a\beta C)+(a\beta\gamma)\ ;\ (a\gamma)=(a\beta\gamma)=(aB\gamma)+(a\beta\gamma)\ ;\ (\beta\gamma)=(A\beta\gamma)+(a\beta\gamma)$

Substituting the values, we get,

$(a)=12000-977=11023;(\beta)=12000-1185=10815$
$(\gamma)=12000-596=11404.$
$(AB\gamma)=453-127=326;(A\beta C)=284-127=157$
$(aBC)=250-127=123$
$(A\beta\gamma)=977-453-157=367;(aB\gamma)=1185-453-123=609$
$(a\beta C)=596-250-157=189;(A\beta)=157+367=524$
$(aB)=123+609=732;(A\gamma)=326+367=693$
$(\alpha C)=123+189=312;(B\gamma)=326+609=935$
$(\beta C)=157+189=346;(\alpha\beta\gamma)=10815-524=189=10{,}102$
$(a\beta)=189+10102=10291;(\alpha\gamma)=609+10102=10711$
$(\beta\gamma)=367+10102=10469$

PROBLEM 9.4. *A number of labourers in a factory were examined for the presence or absence of certain defects of which three chief descriptions were noted :*

A—Physical weakness ; B—Nerve signs ; C—Mental dullness.

Given the following ultimate frequencies, find the frequencies of the positive classes including the whole number of observations N :—

$(ABC)=75;(\alpha BC)=98\ ;\ (AB\gamma)=310;(\alpha B\gamma)=702$
$(A\beta C)=106;(\alpha\beta C)=74\ :\ (A\beta\gamma)=489;(\alpha\beta\gamma)=8415$

SOLUTION. The positive classes including the whole number of observations N, are :

N	(AB)
(A)	(AC)
(B)	(BC)
(C)	(ABC)

Now,

$N=(ABC)+(AB\gamma)+(A\beta C)+(\alpha BC)+(A\beta\gamma)+(\alpha B\gamma)+(\alpha\beta C)+(\alpha\beta\gamma)$

$(A)=(ABC)+(AB\gamma)+(A\beta C)+(A\beta\gamma)$

$(B)=(ABC)+(AB\gamma)+(\alpha BC)+(\alpha B\gamma)$

$(C)=(ABC)+(\alpha BC)+(A\beta C)+(\alpha\beta C)$

$(AB)=(ABC)+(AB\gamma)$; $(AC)=(ABC)+(A\beta C)$

$(BC)=(ABC)+(\alpha BC)$; $(ABC)=75$ (given)

Substituting the given values, we get,

$N=75+310+106+98+489+702+74+8415=10269$

$(A)=75+310+106+489=980$; $(B)=75+310+98+702=1185$

$(C)=75+98+106+74=353$; $(AB)=75+310=385$;

$(AC)=75+106=181$; $(BC)=75+98=173$; $(ABC)=75$

PROBLEM 9.5. *Given the following frequencies, find out the frequencies of the positive and the ultimate classes.*

$N=2{,}90{,}02{,}525$

$(A)=23{,}467$

$(B)=14{,}192$

$(C)=97{,}383$

$(AB\gamma)=82$

$(A\beta C)=380$

$(\alpha BC)=500$

$(ABC)=25$

SOLUTION. The classes for which frequencies have to be found out are :—

(AB)	$(A\beta\gamma)$
(AC)	$(\alpha B\gamma)$
(BC)	$(\alpha\beta C)$
	$(\alpha\beta\gamma)$

Now,

$(AB)=(ABC)+(AB\gamma)$

$(AC)=(ABC)+(A\beta C)$

$(BC)=(ABC)+(\alpha BC)$

$(A\beta\gamma)=(A\beta)-(A\beta C)=(A)-(AB)-(A\beta C)$

$(\alpha B\gamma)=(\alpha B)-(\alpha BC)=(B)-(AB)-(\alpha BC)$

$(\alpha\beta C)=(\beta C)-(A\beta C)=(C)-(BC)-(A\beta C)$

$(\alpha\beta\gamma)=(\alpha\beta)-(\alpha\beta C)$

$=(\beta)-(\alpha\beta C)=(\alpha\beta C)=N-(B)-(A)+(AB)-(\alpha\beta C)$

Substituting the values, we get,

$(AB)=25+82=107.$

$(AC)=25+380=405.$

$(BC)=25+500=525.$

$(A\beta\gamma)=23{,}467-107-380=22{,}980.$

$(\alpha B\gamma)=14{,}192-107-500=13{,}585.$

$(\alpha\beta C)=97{,}383-525-380=96{,}478.$

$(\alpha\beta\gamma)=29002{,}525-14192-23467+107-96{,}478.$

$=2{,}88{,}68{,}495.$

PROBLEM 9.6. *Measurements are made on a thousand husbands and a thousand wives. If the measurements of the husbands exceed the measurements of the wives in 789 cases for one measurement, 741 cases for another, and in 690 cases for both measurements, in how many cases will both measurements on the wife exceed the measurement on the husband ?*

SOLUTION. Let A stand for the cases in which measurements of husbands exceed those of wives in one measurement and B, those in the other measurement.

Then, the data given are :

$N=1000$

$(A)=789$; $(B)=741$: $(AB)=690$

We have to find the frequencies of $(\alpha\beta)$, which are equal to—

$N-(A)-(B)+(AB)$

Substituting the above values, we get,

$\alpha\beta=1000-789-741+690=160.$

Thus the number of cases in which both measurements on the wife exceed the measurements on the husband is 160.

PROBLEM 9.7. *In a Girls' High School there were 200 students. Their results in the quarterly, half-yearly and the annual examinations were as follows :*

80 passed the quarterly examination,
75 passed the half-yearly examination,
and 96 passed the annual examination.
25 passed all three, 46 failed all three.
29 passed the first two and failed in the annual examination.
42 failed the first two but passed the annual examination.
Find how many students passed at least two examinations.

SOLUTION. Denoting :

Success in quarterly examination by (A), and failure by (α)
Success in half-yearly examination, ,, (B), ,, ,, ,, ,, (β)
Success in annual examination ,, (C), ,, ,, ,, ,, (γ)

The given data are reduced to

$N=200$;$(ABC)=25$

$(A)=80$;$(\alpha\beta\gamma)=46$; $(B)=75$; $(AB\gamma)=29$; $(C)=96$;$(\alpha\beta C)=42$

We have to find out the value of

$(\alpha BC)+(A\beta C)+(AB\gamma)+(ABC)$

Now, $(\alpha BC)+(A\beta C)+(AB\gamma)+\alpha\beta C)=(C)$

$\therefore$ $(\alpha BC)+(A\beta C)=(C)-(ABC)-(\alpha\beta C)=96-25-42=29.$

$\therefore$ $(\alpha BC)+(A\beta C)+(AB\gamma)+(ABC)=29+29+25=83.$

Thus, the number of students who passed at least two examinations is 83.

PROBLEM 9.8. *In a free vote in the House of Commons 600 members voted. 300 government members representing English constituencies (including Welsh) voted in favour of the motion, 25 opposition members representing Scottish constituencies voted against the motion. The government majority among those who voted was 96. 135 of the members voting represented Scottish constituencies; 18 government members voted against the motion. 102 Scottish members voted in favour of the motion. The motion was carried by 310 votes. Analyse the voting according to the nationality of the constituencies and party.*

SOLUTION. Denoting the government and opposition parties by A and α respectively, voting for and against the motion by B and β, and English and Scottish members by C and γ respectively, our data are :

$N=600$ ... ... (*i*)
$(ABC)=300$ ... ... (*ii*)
$(\alpha\beta\gamma)=25$ ... ... (*iii*)
$(A)-(\alpha)=96$ ... ... (*iv*)
$(\gamma)=135$ ... ... (*v*)
$(A\beta)=18$ ... ... (*vi*)
$(B\gamma)=102$ ... ... (*vii*)
$(B)-(\beta)=310$ ... ... (*viii*)

The ultimate class-frequencies have to be found out.

From (a)

$(A)+(a)=600$ ... ... (*ix*)
Adding (*i*) and (*iv*), we have
$(A)=348$ ... ... (*x*)
Similarly from, (*i*)
$(B)+(\beta)=600$ ... ... (*xi*)
Adding (*xi*) and (*viii*), we have
$(B)=455$ ... ... (*xii*)
Now $(C)=N-(\gamma)\ =600-135=465$ ... (*xiii*)
and $(AB)=(A)-(A\beta)=348-18=330$ ... (*xiv*)
and $(BC)=(B)-(B\gamma)=455-102=353$ ... (*xv*)

we also have,

$$\alpha\beta\gamma=(\beta\gamma)-(A\beta\gamma)=(\gamma)-(B\gamma)-\{(A\beta)-(A\beta C)\}$$
$$=(\gamma)-(B\gamma)-(A+(AB)+(AC)-(ABC)$$

Substituting the known values on the right and the value of $(\alpha\beta\gamma)$, we have :

$$25=135-102-348+330+AC-300$$

or $AC=310$ ... ... (*xvi*)

Now, to calculate the frequencies of the ultimate classes, we proceed as follows :

$(ABC)=300$ (given)
$(AB\gamma)=(AB)-(ABC)=330-300=30$
$(A\beta C)=(AC)-(ABC)=310-300=10$
$(\alpha BC)=(BC)-(ABC)=353-300=53$
$(A\beta\gamma)=(A\beta)-(A\beta C)=\ 18-\ 10=\ 8$
$(\alpha B\gamma)=(\alpha B)-(\alpha BC)=(B)-(AB)-(\alpha BC)=455-330-53=72.$
$(\alpha\beta C)=(\beta C)-(A\beta C)$
$\qquad=(C)\ -(BC)-(A\beta C)=465-353-10=102.$
$(\alpha\beta\gamma)=25$ (given)

PROBLEM 9.9. *In a war between White and Red forces there are more Red soldiers than White ; there are more armed whites than unarmed Reds ; there are fewer armed Reds with ammunition than unarmed Whites without ammunition. Show that there are more armed Reds without ammunition than unarmed Whites with ammunition.*

SOLUTION. Writing :

A to denote the property of being a White soldier and hence α to denote the property of being a Red soldier ;

Writing

B and β to denote armed and unarmed respectively ; and writing C and γ to denote the possession and non-possession of ammunition, respectively, our data are:

$(\alpha) > (A)$ ... ... (*i*)
$(AB) > (\alpha\beta)$ ... ... (*ii*)
$(A\beta\gamma) > (\alpha BC)$ ... ... (*iii*)

We have to show that

$(\alpha B\gamma) > (A\beta C)$

From (*i*), considering the dichotomy of each side according to B, we have :

$(\alpha B)+(\alpha\beta)>(AB)+(A\beta)$

Substituting for AB from (*ii*) in this inequality.

$(\alpha B)+(\alpha\beta)>(\alpha\beta)+(A\beta)$

and hence

$(\alpha B)>(A\beta)$

From this, considering the dichotomy of each side according to (*iii*), we have

$(\alpha BC)+(\alpha B\gamma)>(A\beta C)+(A\beta\gamma)$

or $(\alpha BC)+(\alpha B\gamma)>(A\beta C)+(\alpha BC)$

Hence $(\alpha B\gamma)>(A\beta C)$

Thus, it has been proved that there are more armed Reds without ammunition than unarmed Whites with ammunition.

PROBLEM 9.10. *Show that for n attributes A, B, C, ... M,*

$(ABC)\ldots M)\geqslant\{(A)+(B)+(C)+\ldots\ldots+(M)\}-(n-1)\ N.$

Where N is the total frequency.

SOLUTION. To prove the above, we proceed as follows :—

$$(\alpha\beta)=(1-A)\,(1-B),\ N$$
$$=N-(A)-(B)+(AB)$$

As $(\alpha\beta)$ cannot have a negative value.

$AB\geqslant(A)+(B)-N.$

Apply the above inequality to the universe of C,

$$\text{Then } ABC\geqslant(AC)+(BC)-(C)$$
$$\geqslant(A)+(C)-N+(B)+(C)-N-(C)$$
$$\geqslant(A)+(B)+(C)-2N.$$

Apply the above to the universe of D

$$\text{Then } (ABCD)\geqslant(AD)+(BD)+(CD)-2(D)$$
$$\geqslant(A)+(D)-(N)+(B)+(D)-N+(C)+(D)-N-2(D)$$
$$\geqslant(A)+(B)+(C)+(D)-3N.$$

Therefore $(ABC\ldots M)\geqslant\{(A)+(B)+(C)+\ldots\ldots+(M)\}-(n-1)N.$

PROBLEM 9.11. *In a very hotly fought battle*
70% at least of the combatants lost an eye.
75% at least an ear.
80% at least a leg.
85% at least an arm.

What percentage at least lost all four ?

SOLUTION. Denoting the combatants who lost an eye by A ; those who lost an ear by B ; those who lost a leg by C ; and those who lost an arm by D and the total number of combatants by N.

$(A)=.70\text{ N}\ ;\ (B)=.75\text{ N}\ ;\ (C)=.80\text{ N}\ ;\ (D)=.85\text{ N}\ ;$

We have to find the least value of (ABCD).

Now,

$(ABCD)\geqslant\{(A)+(B)+(C)+(D)\}-(n-1)\text{ N}$; where n represents the no. of attributes.

$$\begin{aligned}\text{Or } (ABCD) &\geqslant .70\text{ N}+.75\text{ N}+.80\text{ N}+.85\text{ N}-3\text{N}.\\ &\geqslant 3.10\text{ N}-3\text{N}\\ &\geqslant .10\text{N}\end{aligned}$$

Hence, the least possible value of (ABCD) is 10%

Thus, 10% of the combatants at least, lost all four.

PROBLEM 9.12. *A universe consists of three attributes each of which is divisible into two parts. What are the different class frequencies obtainable ?*

At a competitive examination at which 600 graduates appeared, boys outnumbered girls by 96, Those qualifying for interview exceeded in number than those failing to qualify by 310. The number of science graduate boys interviewed was 300 while among the Arts graduate girls, there were 25 who failed to qualify for interview. Altogether there were only 135 Arts graduates and 33 among them failed to qualify. Boys who failed to qualify numbered 18. Find (a) the number of boys who qualified for interview, (b) the total number of science graduate boys appearing and (c) the number of science graduate girls who qualified.

SOLUTION. Suppose A stands for Boys, α, for Girls.

B stands for those who qualified for interview and β for those who failed t qualify.

C stands for science graduates and γ for arts graduates.

The given data would be—

$N=600\ ;\ (A)-(\alpha)=96\ ;\ (B)-(\beta)=310$
$(ABC)=300\ ;\ (\alpha\beta\gamma)=25\ ;\ (\gamma)=135\ ;\ (\beta\gamma)=33\ (A\beta)=18.$

We have to find out the values of (AB), (AC) and (αBC)

Now $(A)-(\alpha)=96.$

$(A)+(\alpha)=600$

$(2A)\quad =696 \text{ or } (A)=348$

Similarly $(B)=455$ and $C=(N-\gamma)=465.$

$(AB)=(A)-(A\beta)=348-18=330.$

$(AC)=(ABC)+(A\beta C)$

$=(ABC)+(A\beta-A\beta\gamma)=(ABC+\{(A\beta)-(B\gamma-\alpha\beta\gamma)\}$

$=(ABC)+(A\beta)-(\beta\gamma)+(\alpha\beta\gamma)=300+18-33+25=310.$

$(\alpha BC)=(BC)-(ABC)=(C)-(\beta C)-(ABC)=[(C)-\{(\beta)-(\beta\gamma)\}-(ABC)]$

$=(C)-(\beta)+(\beta\gamma)-(ABC)=465-145+33-300=53.$

(ASSOCIATION)

PROBLEM 9.13. *Show as briefly as possible, whether A and B are independen positively associated or negatively associated in each of the following cases :—*

(*i*) $N=50000$, $(A)=23500$, $(B)=31000$, $(AB)=16000$,
(*ii*) $A=4900$, $(AB)=2940$, $(\alpha)=5700$, $(\alpha B)=3800$,
(*iii*) $(AB)=2560$, $(\alpha B)=7680$, $(A\beta)=480$, $(\alpha\beta)=1440$,

SOLUTION. (*i*) Expected frequencies of $AB=\dfrac{A\times B}{N}=$

$$\frac{23500\times 31000}{50000}=14570$$

Actual observation of AB is 16000, which is more than the expectation. Hence there is a positive association between A and B.

(*ii*) Proportion of B in A's, *i.e.*, $AB/A = \dfrac{2940}{4900}=\dfrac{3}{5}$

and ,, ,, ,, B in α's *i.e.*, $\alpha B/\alpha = \dfrac{3800}{5700}=\dfrac{2}{3}$

Thus $\dfrac{AB}{A}<\dfrac{\alpha B}{\alpha}$

Therefore, there is a negative association between A and B.

(*iii*) $(AB)\times(\alpha\beta)=2560\times 1440=3686400$
and $(A\beta)\times(\alpha B)=480\times 7680=3686400$
$(AB)\times(\alpha\beta)=(A\beta)\times(\alpha B)$
Thus, A and B are independent.

PROBLEM 9.14. *Investigate the association between darkness of eye-colour in father and son from the following data :*

Father with dark eyes and sons with dark eyes=500
Father with dark eyes and sons with not dark eyes=790
Father with not dark eyes and sons with dark eyes=890
Father with not dark eyes and sons with not dark eyes=7820

SOLUTION. Denoting

Fathers with dark eyes by A and fathers with not dark eyes by a and sons with dark eyes by B and sons with not dark eyes by β

Our data are :

$(AB)=500$, $(aB)=890$
$(A\beta)=790$, $(a\beta)=7820$

Now,

Percentage of dark-eyed amongst the sons of dark-eyed fathers i.e. $\dfrac{AB}{A}\times 100$

$$\text{or } \frac{AB}{AB+A\beta}\times 100=\frac{500}{500+790}\times 100=\frac{50000}{1290}$$

$=39\%$ approx.

Similarly percentage of dark-eyed amongst the

sons of not dark-eyed fathers, i.e., $\frac{\alpha B}{\alpha}\times 100=\frac{890}{390+7820}\times 100$

$$=\frac{89000}{8710}=10\% \text{ approx.}$$

A comparison of two percentages reveals clearly that there is a positive association between darkness of eye-colour in father and in son.

PROBLEM 9.15. *Can vaccination be regarded as a preventive measure for small-pox from the data given below ?*

Of 1,482 persons in a locality exposed to small-pox, 368 in all were attacked. Of 1,482, persons, 343 had been vaccinated and of these only 35 were attacked.

SOLUTION. Let A be the attribute of vaccination,

α, that of non-vaccination
B, that of exemption from attack of small-pox
and β, that of attack of small-pox.

Then, our data are :

$N=1482; (A)=343; (\beta)=368; (A\beta=35;$

From the above $(A)=343$
$(AB)=(A)-(A\beta)=343-25=308$
$(\alpha)=N-(A)=1482-343=1139$
$(\alpha B)=(B)-(AB)=N-(\beta)-(AB)=1482-368-308=806$
Now, percentage of vaccinated who were not

attacked, *i.e.*. $\frac{AB}{A}\times 100=\frac{308}{343}\times 100$

$$=90\%$$

and percentage of not vaccinated who

were not attacked, *i.e.* $\frac{\alpha B}{\alpha}\times 100=\frac{806}{1139}\times 100=71\%$

A comparison of the two percentages brings out the fact that vaccination and exemption from attack are positively associated. Hence, vaccination may be regarded as preventive measure for small-pox.

PROBLEM 9.16 *In an anti-malarial campaign in a certain area, quinine was administered to 812 persons out of a total population of 3248. The number of fever cases is shown below :—*

Treatment	*Fever*	*No Fever*
Quinine	*20*	*792*
No-quinine	*220*	*2216*

Discuss the usefulness of quinine in checking malaria.

SOLUTION. Denoting.

Quinine treatment by A and no quinine treatment by α
and no attack of fever by B and attack of fever by β
our data are:

$(AB)=792, (\alpha B)=2216$
$(A\beta=20, (\alpha\beta)=220.$

From the above,

$(A)=(AB)+(A\beta)=792+20=812$

From the above,

$(a)=(aB)+(a\beta)=2216+220=2436$

Now,

percentage quinine treatment cases which

were attacked, *i.e.* $\frac{A\beta}{A}\times 100=\frac{20}{812}\underset{1}{\times}100=2.5\%$

and percentage of no quinine treatment cases which

were attacked, *i.e* $\frac{a\beta}{a}\times 100=\frac{220}{2439}\times 100=9\%$

A comparison of the two percentages brings out the fact that quinine tr ment and attack from fever are negatively associated. From the above, it may concluded that quinine is effective in checking malaria.

PROBLEM 9.17. *Can it be concluded from the data given in the follow table that bottle-feeding is conducive to malocclusion of teeth ?*

Malocclusion of the Teeth in Infants

	Normal Teeth	*Malocclus*
Breast-fed	*9*	*11*
Bottle-fed	*1*	*21*

SOLUTION. Denoting

Bottle-feeding by A; and Breast-feeding by a
Malocclusion by B; and Normal teeth by β.
the data are :

$(AB)=21; (aB)=11; (a\beta)=9$

From the above,

$(A)=(AB)+A\beta)=21+1=22; (a)=(aB)+(a\beta)=11+9=20.$

Now, percentage of bottle-feeding cases leading

to malocclusion, *i.e.,* $\frac{AB}{A}\times 100=95.5\%$

and percentage of breast-feeding cases leading

to malocclusion, *i.e.* $\frac{aB}{a}\times 100=\frac{11}{20}\times 100=55\%$

A comparison of the two percentages brings out clearly the fact that bo feeding is conducive to malocclusion of teeth.

PROBLEM 9.18. *Do you find any association between the tempers of broth and sisters from the following data :—*

Good-natured brothers and good-natured sisters	*1230*
Good-natured brothers and sullen sisters	*850*
Sullen brothers and good-natured sisters	*530*
Sullen brothers and sullen sisters	*980*

SOLUTION. Denoting.

Good-natured brothers by A and sullen brothers by a

Good natured sisters by B and sullen sisters by β.

the given data are $(AB)=1230$; $(A\beta)=850$; $(a\beta)=530$; $(a\beta)=980$

From the above,

$$(A)=(AB)+(A\beta)=1230+850=2080$$
$$(a)=(aB)+(a\beta)=530+980=1510$$

Now,

Percentage of good-natured sisters amongst the sisters of the good-natured brothers, that is $\frac{AB}{A}\times 100=\frac{1230}{2080}\times 100=59\%$.

and Percentage of good-natured sisters amongst the sisters of the sullen brothers, that is $\frac{aB}{a}\times 100=\frac{530}{1510}\times 100=35\%$

A comparison of the two percentages brings out clearly the fact that the tempers of the brothers and the sisters are positively associated.

PROBLEM 9.19. *Explain the method of finding association between two attributes. Out of 70 thousand of literates in a particular district of India, number of criminals was 500. Out of 930 thousand of illiterates in the same district, number of criminals was 15 thousand. On the basis of these figures, do you find any association between illiteracy and criminality ?*

SOLUTION. Denoting :

Illiterates by A ; literates by a ;

and criminals by B ; and non-criminals by β the data are :

$(A)=930{,}000$ $(AB)=15{,}000$

$(a)=70{,}000$ $(aB)=500$

Now,

the percentage of literates who are criminals ;

that is, $\frac{AB}{A}\times 100=\frac{15{,}000}{930{,}000}\times 100=1.6\%$

and the percentage of literates who are criminals,

that is, $\frac{aB}{a}\times 100=\frac{500}{70{,}000}\times 100=.7\%$

A comparison of the above two percentages brings out clearly the fact that illiteracy and criminality are positively associated.

PROBLEM 9.20. (*a*) *Write a short note on the use of Co-efficient of Association in analysing economic statistics.*

(*b*) *From the figures given in the following table, compare the association between literacy and unemployment in rural and urban areas, and give reasons for the difference, if any :—*

	Urban	*Rural*
Total Adult Males	*25 lakhs*	*200 lakhs*
Literate Males	*10 lakhs*	*40 lakhs*
Unemployed Males	*5 lakhs*	*12 lakhs*
Literate and unemployed Males	*3 lakhs*	*4 lakhs*

SOLUTION. Denoting :

Literate males in lakhs by A, and unemployed males in lakhs by B, the given data are :

	Urban	Rural
N	25	200
(A)	10	40
(B)	5	12
(AB)	3	4

Urban

From the above,

$(AB)=3$
$(A\beta)=(A)-(AB)=10-3=7$
$(aB)=(B)-(AB)=5-3=2$
$(a\beta)=(a)-(aB)=N-(A)-(aB)=25-10-2=13$

Substituting these values in the Yule's formula for co-efficient of association :

$$Q=\frac{(AB)\,(a\beta)-(A\beta)\,(aB)}{(AB)\,(a\beta)+(A\beta)\,(aB)}$$

We get,

$$Q=\frac{(3)\,(13)-(7)\,(2)}{(3)\,(13)+(7)\,(2)}=\frac{39-14}{39+14}=\frac{25}{53}=+.47$$

Rural :

From the above,

$(AB)\ =4$
$(A\beta)\ =\ (A)-(AB)=40-4\ =36$
$(aB)\ =\ (B)-(AB)=12-4\ =8$
$(a\beta)\ =N-(A)-(aB)\ =200-40-8=152.$

Substituting these values in the Yule's formula :

$$Q=\frac{(AB)\,(a\beta)-(A\beta)\,(aB)}{(AB)\,(a\beta)+(A\beta)\,(aB)}$$

Where Q represents the co-efficient of association

We get,

$$Q=\frac{(4\times152)-(36\times8)}{(4\times152)+(36\times8)}=\frac{608-288}{608+288}=\frac{320}{896}=+.357.$$

Thus, there is a positive association between literacy and unemployment both in rural and urban areas. It is, however, less in the case of rural as compared to urban areas.

PROBLEM 9.21. *The following table gives the number of literates and criminals in three cities of U.P.*

	Meerut	*Bareilley*	*Saharanpur*
Total number (in thousands)	*244*	*184*	*230*
Literates (in thousands)	*40*	*47*	*33*

Literate criminals (in hund-eds)	*3*	*2*	*2*
Illiterate criminals (in hundreds)	*40*	*20*	*24*

Compare the degree of association between criminality and illiteracy in each of the above three towns.

SOLUTION. Denoting :

Illiterates (in hundreds) by A and literates (in hundreds) by a and criminals (in hundreds) by B, and non-criminals (in hundreds) by β the given data are,

	Meerut	Bareilley	Saharanpur
N	2440	1840	2300
(a)	400	470	330
(aB)	3	2	2
(AB)	40	20	24

From the above,

Second-order frequencies of the data may be put down as follows :

(AB)	$= 40$	20	24
(aB)	$= 3$	2	2
$(A\beta) = N - (a) - (AB)$	$= 2000$	1350	1946
$(a\beta) \quad (a) - (aB)$	$= 397$	468	328

Substituting the above values in the Yule's formula

$$Q = \frac{(AB)\,(a\beta) - (A\beta)\,(aB)}{(AB)\,(a\beta) + (A\beta)\,(aB)}$$

Where Q represents the co-efficient of association,

We get,

$$Q \text{ Meerut} = \frac{(40 \times 397) - (2000 \times 3)}{(40 \times 397) + (2000 \times 3)} = \frac{9880}{21880} = +.45 \text{ approx.}$$

$$Q \text{ Bareilley} = \frac{(20 \times 468) - (1350 \times 2)}{(20 \times 468) + (1350 \times 2)} = \frac{6660}{12060} = +.55 \text{ approx.}$$

$$Q \text{ Saharanpur} = \frac{(24 \times 328) - (1946 \times 2)}{(24 \times 328) + (1946 \times 2)} = \frac{3980}{11764} = +.34 \text{ approx.}$$

Thus, there is a positive association between criminality and illiteracy in all the three towns. But it is more in Meerut and Bareilley than in Saharanpur. As between Bareilley and Meerut it is more in Bareilley.

PBOBLEM 9.22. *A census revealed the following figures of the blind and the insane in two age-groups in a certain population.*

		Age-group *15—25 years*	*Age-group* *Over 25 years*
Total population	...	*2,70,000*	*1,60,200*
Number of blind	...	*1,000*	*2,000*
Number of insane	...	*6,000*	*1,000*
Number of insane among the blind	...	*19*	*9*

(*a*) *Obtain a measure of the association between blindness and insanity in each of the two age-groups.*

(*b*) *Do you consider that blindness and insanity are associated or disassociated with each other in the two age-groups, or more in one age-group than in other ?*

SOLUTION. Denoting :

Blindness by A, and non-blindness by a, insanity by B, and sanity by β the given data are :

		Age-group 15—25 years	Age-group over 25 years
N	=	2,70,000	1,60,200
(A)	=	1,000	2,000
(B)	=	6,000	1,000
(AB)	=	19	9

From the above the class-frequencies of the second order of the two age-groups may be put down as follows :—

		Age-group 15—25 years	*Age-group* Over 25 years
(AB)	=	19	9
$(A\beta)=(A)-(AB)$	=	881	1,991
$(aB)=(B)-(AB)$	=	5,981	991
$(a\beta)=(a)-(aB)=$ $=N-(A)-(aB)$	=	2,63,019	1,57,209

Substituting the above values in the Yule's formula :

$$Q=\frac{(AB)\,(a\beta)-(A\beta)\,(aB)}{(AB)\,(a\beta)+(A\beta)\,(aB)}$$

Where Q represents the co-efficient of association,

We get,

$$\text{Q for 15—25 age-group}=\frac{(19\times263019)-(981\times5981)}{(19\times263019)+(981\times5981)}$$

$$=\frac{4997361-5867361}{4997361+5867361}=\frac{-870000}{10864722}=-.08$$

$$\text{and Q for over 25 age-group}=\frac{(9\times157209)-(1991\times991)}{(9\times157209)+(1991\times991)}$$

$$=\frac{1414881-1973081}{1414881+1973081}=\frac{-558200}{3387962}=-.16$$

Thus, there is a disassociation between blindness and insanity in both the age-groups. But the degree of disassociation is more in the 'over 25 years age-group' than in the '15—25 years age-group.'

PROBLEM 9.23. *Calculate the Co-efficient of Association between extravagance in fathers and sons from the following data :—*

Extravagant fathers with extravagant sons	...	327
Extravagant fathers with miserly sons	...	545
Miserly fathers with extravagant sons	...	741
Miserly fathers with miserly sons	...	235

SOLUTION. Denoting :

Extravagant fathers by A, and miserly fathers by a, extravagant sons by B and miserly sons by β, the given data are :

$(AB)=327$; $(A\beta)=545$; $(aB)=741$; $(a\beta)=235$.

Substituting the above values in the Yule's formula for co-efficient of association,

$$Q=\frac{(AB)\,(\alpha\beta)-(A\beta)\,(\alpha B)}{(AB)\,(\alpha\beta)+(A\beta)\,(\alpha B)}$$

Where Q represents the co-efficient of association,

We get,

$$Q=\frac{(327\times 235)-(545\times 741)}{(327\times 235)+(545\times 741)}=\frac{76845-403845}{76845+403845}=\frac{-327000}{480690}=-.68$$

Thus, there is a negative association between extravagance in fathers and sons. The co-efficient of association between the two is —.68.

PROBLEM 9.24. *The male population of a certain state in India is 331 lakhs. The number of literate males is 66 lakhs and the number of male criminals is 33 thousand. If the number of literate male criminals is 6 thousand, calculate the co-efficient of association between literacy and criminality in this state.*

SOLUTION. Denoting literate by 'A', illiterate by '*a*' and criminal by 'B' and not-criminal by β.

The data are : $N = 3{,}31{,}00{,}000$
$A = 66{,}00{,}000$
$B = 33{,}000$
$AB = 6{,}000$

from the above,

$(AB)=6{,}000$

$(A\beta)=(A)-(AB)=66{,}00{,}000-6000=65{,}94{,}000$

$(\alpha B)=(B)-(AB)=33{,}000-6000=27{,}000$

$(\alpha\beta)-(\alpha)-(\alpha B)$

$=N-(A)-(\alpha B)$

$=3{,}31{,}00{,}000-66{,}00{,}000-27{,}000$

$=264{,}73{,}000$

Substituting these values in the Yule's formula for co-efficient of association :—

$$Q=\frac{(AB)\,(\alpha B)-(A\beta)\,(\alpha B)}{(AB)\,(\alpha B)+(A\beta)\,(\alpha B)}$$

We get,

$$Q=\frac{(6000\times 2{,}64{,}73{,}000)-(65{,}94{,}000\times 27{,}000)}{(6000\times 2{,}64{,}73{,}000)+(65{,}94{,}000\times 27{,}000)}=\frac{17{,}20{,}00{,}00{,}000}{3{,}34{,}87{,}60{,}000{,}00}$$

$$=-.056$$

Thus, there is hardly any association between literacy and criminality.

PROBLEM 9.25. *In an investigation into the Health and Nutrition of certain children of Kanpur between the ages of one and five years, two groups of children were compared one belonging to the professional classes, 125 in number, and the other belonging to the labouring classes, 124 in number. The following results were deduced :—*

	Poor children per cent	*Well-to-do children per cent*
Below normal weight	75	23
Above normal weight	5	42

Find the co-efficient of association between the weight of the children and social status.

SOLUTION. Denoting :

Poor children by A, and well-to-do children by a and below normal weigh B, and above normal weight by β the given data are :

$(AB)=75$; $(A\beta)=5$; $(aB)=23$; $(a\beta)=42$

Substituting the above values in the Yule's formula for co-efficient of asso tion.

$$Q=\frac{(AB)\,(a\beta)-(A\beta)\,(aB)}{(AB)\,(a\beta)+(A\beta)\,(aB)}$$

Where Q represents the co-efficient of association.
We get,

$$Q=\frac{(75\times42)-(5\times23)}{(75\times42)+(5\times23)}=\frac{3150-115}{3150+115}=\frac{3025}{3265}=+.929$$

Thus, there is a positive association between the weight of the children and th social status. The co-efficient of association between the two is $+.929$.

PROBLEM 9.26. *In December 1974 there was an outbreak of cholera in a j in Bombay. Of 127 persons who were uninoculated, 10 contracted cholera, 6 of th dying. Of 147 persons who were inoculated, 3 contracted cholera and there were deaths.*

Trace the association between [a] inoculation and contracting the disease [b] in oculation and mortality among persons who have contracted the disease.

SOLUTION. (*a*) Denoting,

Inoculation by A, absence of it by α, infection by B and absence of it by β. Th given data are,

$(A)=147; (\alpha)=127.$
$(\alpha B)=10; (AB)=3.$
$(A\beta)=(A)-(AB)=(147-3)=144.$
$(\alpha\beta)=(\alpha)-(\alpha B)=(127-10)=117.$

Substituting these values in Yule's formula,

$$Q=\frac{(AB)\,(\alpha\beta)-(A\beta)\,(\alpha B)}{(AB)\,(\alpha\beta)+(A\beta)\,(\alpha B)}$$ where

Q is the co-efficient of association, we get,

$$Q=\frac{(3\times117)-(144\times10)}{(3\times117)+(144\times10)}=\frac{351-1440}{351+1440}=\frac{-1089}{1791}=-.6$$

Thus the co-efficient of association, between inoculation and contracting th disease is—.6.

(*b*) If A and α denote inoculation and its absence respectively, amongst thos who contracted the disease and B and β represent death and recovery respectivel the given data are :

$(A)=3$; $(\alpha)=10.$
$(AB)=0$; $(\alpha B)=6.$

The values of

$(A\beta)=(A)-(AB)=3-0=3; (\alpha\beta)=(\alpha)-(\alpha B)=10-6=4.$

Substituting the values in Yule's formula of co-efficient of association, we get,

$$Q=\frac{(0\times4)-(3\times6)}{(0\times4)+(3\times6)}=\frac{-18}{18}=-1$$

Thus the co-efficient of association between inoculation and mortality among persons who have contracted the disease is -1.

PROBLEM 9.27. *The following table shows the distribution of the temper in pairs of sisters in an exhaustive school enquiry :—*

Second Sister	*First Sister* *Good-natured*	*Sullen*	*Total*
Good-natured	1040	180	1220
Sullen	160	120	280
Total	1200	300	1500

Trace the association, if any, in the distribution of tempers in first sister and second sister.

SOLUTION. Denoting second sister good-natured by A and sullen by α, and first sister good-natured by B and sullen by β, the given data are :

$AB=1040,\ A\beta=180,\ \alpha B=160,\ \alpha\beta=120.$

Substituting the above values in Yule's formula of co-efficient of association.

$$Q=\frac{(AB)\,(\alpha\beta)-(A\beta)(\alpha B)}{(AB)\,(\alpha\beta)+(A\beta)(\alpha B)}$$

We get,

$$Q=\frac{(1040\times120)-(180\times160)}{(1040\times120)+(180\times160)}=\frac{124800-28800}{124800+28800}=\frac{96{,}000}{1{,}53{,}600}=.625.$$

Thus the co-efficient of association between the temperament of first an second sister is .625.

PROBLEM 9.28. *Find out the co-efficient of association between the type o college training and success in teaching from the following table :—*

Institution		*Successful*	*Unsuccessful*	*Total*
Teachers' College	...	58	42	100
University	...	49	51	100
Total		107	93	200

SOLUTION. Denoting.

Teachers' College by A, University by α, successful by B and unsuccessful β, the given data :

$(AB)=58;(A\beta)=42;(\alpha B)=49;(\alpha\beta)=51$

Substituting the above values in Yule's formula for co-efficient of associatio

$$Q=\frac{(AB)\,(\alpha\beta)-(A\beta)\,(\alpha B)}{(AB)\,(\alpha\beta)+(A\beta)\,(\alpha B)}$$

$$Q=\frac{(58\times51)-(42\times49)}{(58\times51)+(42\times(49)}=\frac{2958-2058}{2958+2058}=\frac{900}{5016}=.18.$$

Thus, the desired co-efficient of association is—.18.

PROBLEM 9.29. *What do you understand by contingency ? Calculate the co-efficient of association between relief recipients and type of relief from the following contingency table :—*

Type of relief	*Race*		*Total*
	Negro	*White*	
Work	*72*	*376*	*448*
Direct	*154*	*250*	*404*
Total	*226*	*626*	*852*

SOLUTION. Denoting.

Work by A, Direct by a, Negro by B and White by β, the given data are $AB=72$; $A\beta=376$; $\alpha B=154$; $\alpha\beta=250$;

Substituting the above values in Yule's formula of co-efficient of association.

$$Q=\frac{(AB)\,(\alpha\beta)-(A\beta)\,(\alpha B)}{(AB)\,(\alpha\beta)+(A\beta)\,(\alpha B)}$$

$$Q=\frac{(72\times250)-376\times154)}{(72\times250)+(376\times154)}=\frac{18000-57904}{18000+57904}=\frac{-39904}{75904}=-.53.$$

Thus the desired co-efficient of association is—.53.

PROBLEM 9.30. *When are two attributes said to be (a) independent, (b) positively associated, (c) negatively associated ? What are the conditions to be satisfied by class-frequencies in each of the above cases ?*

Investigate the association between darkness of eye-colour in father and son from the following data :—

Combination	*frequenc*
Fathers with dark eyes and sons with dark eyes	*50*
Fathers with dark eyes and sons with not dark eyes	*79*
Fathers with not dark eyes and sons with dark eyes	*89*
Fathers with not dark eyes and sons with not-dark eyes	*782*

What would have been the frequency of fathers with dark eyes and sons with dark eyes, for the same total number, had there been complete independence ?

SOLUTION. Denoting.

Father with dark eyes by 'A' and father with not dark eyes by 'a' sons with dark eyes by 'B' and sons with not dark eyes by 'β' the given data are :—

$(AB)=50$; $(A\beta)=79$; $(aB)=89$; $(a\beta)=782$

Substituting the above values in the Yule's formula for co-efficient of association.

$$Q=\frac{(AB)\,(a\beta)-(A\beta)\,(aB)}{(AB)\,(a\beta)+(A\beta)\,(aB)}$$

We get,

$$Q=\frac{(50\times782)-(79\times89)}{(50\times782)\times(79\times89)}=\frac{39100-7031}{39100+7031}=\frac{32069}{46131}=+.69$$

Thus, there is positive association between darkness of eyes of fathers and sons.

If the two attributes were independent the frequency of "fathers with dark eyes and sons with dark eyes" would have been,

$\frac{(A)\times(B)}{N}$; $(A)=(AB)+(A\beta)=50+79=129$

$(B)=(AB)+(aB)=50+89=139$

$N=(AB)+(A\beta)+(aB)+(a\beta)=50+79+89+782=1000$

$$\frac{(A)\times(B)}{N}=\frac{129\times139}{1000}=18$$

PROBLEM 9.31. *Explain clearly what you understand by association and illusory association.*

In a state with a total population of 70,000 adults, 34,000 are males and out of a total of 6,000 graduates, 700 are females. Out of 1200 graduate employees of the state, 200 are females. Is there any sex bias in education among people ? The state holds that no distinction is made in appointment in respect of sex ? How far is their claim substantiated by the data given above ?

SOLUTION. The percentage of males who are graduates is $\frac{5300\times100}{34,000}$ or 16 approx.

The percentage of females who are graduates is $\frac{700\times100}{36000}$ or 2 approx.

These percentages clearly show that there is a sex bias in education. A larger percentage of males than females get education.

If there is no sex distinction in matters of appointment, then out of 1200 graduate employees the number of females should be.

$\frac{700\times1200}{6000}$ or 140.

Their actual number is 200 which shows that there is a bias in favour of females.

REVISIONAL PROBLEMS -9

PROBLEM 9.32. In a population of 400 students, the number of marrie 160. Out of 120 students who failed, 48 belonged to the married group. It is requ ed to find out whether the attributes of marriage and failure are independent, po- tively associated or negatively associated.

[*Ans.* (*i*) *Proportion of marr ied students who failed 30%,*
(*ii*) *Proportion of unmarried students who failed 30%*

Since the two proportions are the same, we can conclude that the attributes of ma riage and failure are independent.

PROBLEM. 9.33. One hundred and seventy six residents of an Indian ci who were interviewed during a sample survey, are classified below according to th smoking and tea drinking habits. Calculate Yule's Co-efficient of Association a comment on its value.

	Smokers	Non-smok
Tea Drinkers	80	66
Non-Tea Drinkers	6	24

[*Ans.* $Q=0.65$

PROBLEM 9.34. If (A)=50 ; (B)=60 and N=100, what would be the va of (AB) if the two attributes A and B are independent ?

[*Ans.* $AB=3$

PROBLEM 9.35. Given N=3,000 ; (B)=720 ; (A)=766 ; (AB)=70. Pre pare 2×2 fold contingency table. Compute Yule's Co-efficient of Association.

[*Ans.* $Q=0.60$

PROBLEM 9.36. The following data relate to literacy and unemployment i a group of 1000 persons. You are required to calculate Yule's Co-efficient of Ass ciation between literacy and unemployment.

Illiterate Unemployed	440
Literate Employed	40
Illiterate Unemployed	360

[*Ans.* $Q=+0.53$

PROBLEM 9.37. In a co-educational institution, out of 1,000 students, 75 were boys. They wrote an examination and it was found that 600 passed, 50 gi had failed. Is there any association between sex and success in the examination ?

[*Ans.* $Q=$

PROBLEM 9.38. From the following ultimate class frequencies, find the fr quencies of the positive and negative classes and the total number of observations :

(AB)=100 ; (*a*B)=80 ; (Aβ)=50 ; (*a*β)=40.

[*Ans.* (*A*)=*150* ; (*B*)=*180* ; (*a*)=*120* ; (β)=*90* ; *N*=*27*

PROBLEM 9.39. From the following data, find out the missing frequencie

(AB)=200 ; (A)=600 ; (B)=1200 ; N=2,000

[*Ans.* (*A*β)—*400* ; (*a*)—*1,400* ; (β)—*800* ; (*aB*)—*1,000* ; (*a*β)=*40*

PROBLEM 9.40. From the following two cases find out, whether the data are consistent or not.

Case I (A)=200 ; (B)=300 ; (AB)=120 ; N=1,000

Case II (A)=200 ; (B)=300 ; (AB)=280 ; N=1,000

[*Ans.* *Case I Consistent* ; *Case II Inconsistent*]

PROBLEM 9.41. Calculate the Co-efficient of Association from the following data :

(1) (A)=60 ; (B)=80 ; (AB)=48 ; N=100
(2) (A)=60 ; (B)=80 ; (AB)=60 ; N=100
(3) (A)=60 ; (B)=80 ; (AB)=40 ; N=100
(4) (A)=60 ; (B)=80 ; (AB)=50 ; N=100

[*Ans.* (*1*) $Q=0$; (2) $Q=+1$; (3) $Q=-1$; (4) $Q=+.25$]

10

Consistence of Data

It is obvious that when Statistics of Attributes are collected no class-frequency should be negative. If any class frequency is negative the data are said to be inconsistent. In order to test the consistence of data we should try to find out if there any class with a negative frequency. If A, B, and C stand for the presence of the attributes and α, β and γ for the absence cf the same the following conditions should be fulfilled otherwise the data would be inconsistent :—

One Attribute

(1)	$A \nless 0$	Otherwise	A	will	be	negative
(2)	$A \ngtr N$	,,	α	,,	,,	,,

Two Attributes

(1)	$AB \nless 0$	,,	AB	,,	,,	,,
(2)	$AB \nless A+B-N$	,,	$\alpha\beta$	,,	,,	,,
(3)	$AB \ngtr A$	,,	$A\beta$	,,	,,	,,
(4)	$AB \ngtr B$	,,	αB	,,	,,	,,

Three Attributes

(1)	$ABC \nless 0$	,,	ABC	,,	,,	,,
(2)	$ABC \nless AB+AC-A$	,,	$A\beta\gamma$	,,	,,	,,
(3)	$ABC \nless AB+BC-B$	,,	$\alpha B\gamma$	,,	,,	,,
(4)	$ABC \nless AC+BC-C$	,,	$\alpha\beta C$	,,	,,	,,
(5)	$ABC \ngtr AB$	,,	$AB\gamma$	,,	,,	,,
(6)	$ABC \ngtr AC$	,,	$A\beta C$	,,	,,	,,
(7)	$ABC \nless BC$	,,	αBC	,,	,,	,,
(8)	$ABC \nless AB+AC+BC-A-B-C+N$	,,	$\alpha\beta\gamma$	,,	,,	,,

In this chapter various types of questions have been solved to find out if the data given are consistent or not.

PROBLEM 10.1. *The following data are given in a report.*

$N=1000$; $(AB)=200$; $(A\beta)=350$; $(\alpha B)=500$.

Show that there must be a misprint or mistake of some sort.

SOLUTION. There is nothing obviously wrong with the figures. Yet they are certainly inconsistent.

$(\alpha\beta)=N-(AB)-(A\beta)-(\alpha B)$,

Substituting the given values, we get,

$(\alpha\beta)=1000-200-350-500=-50$.

It is obvious that no class-frequency, occurring by counting of real attributes can be negative. Hence there must be a misprint or mistake of some sort in the given data.

PROBLEM 10.2. *To investigate the association between eye-colour of husband and eye-colour of wife, the following data are available.*

Husbands with light eyes and wives with not-light eyes	...	*414*
Husbands with non-light eyes and wives with light eyes	...	*260*
Husbands with not-light eyes and wives with not-light eyes	...	*238*
Husbands with light eyes	...	*400*

Do you find any inconsistency in the above data.

SOLUTION. Denoting.

Husbands with light eyes by	A and
Husbands with not-light eyes by	α
Wives with light eyes by	B and
Wives with not-light eyes by	β

the given data are :

$(A\beta)=414$; $(\alpha B)=260$; $(\alpha\beta)=238$; $(A)=400$.

From the above, $(AB)=(A)-(A\beta)=400-414=-14$.

Thus AB has got a negative value. Hence the given data are inconsistent as it is obvious that no class-frequency occurring by counting real attributes can be negative.

PROBLEM 10.3. *In an investigation measurements were made on a thousand husbands and a thousand wives. From the records of the said investigation, it is found that the measurements of the husbands exceeded the measurements of the wives in 800 cases for one measurement, in 700 cases for another and in 420 cases for both measurements.*

Show that the data are inconsistent as they stand.

SOLUTION. Let A be the measurement of husbands exceeding that of the wives in one, and B in the other case.

Then,

$(A)=800$; $(B)=700$; $(AB)=420$; $N=1000$

Now,

$(AB)\nless(A)+(B)-N$: or $(AB)\nless800+700-1000$; or $(AB)\nless500$

But the given value of AB which is 420 is less than the above value. Hence the data are inconsistent.

PROBLEM 10.4. *A labour welfare officer returns the following number of workers observed with certain classes of defects amongst a number of factory workers. A denotes development defects and B low nutrition.*

$N=600$; $(A)=250$; $(\alpha B)=400$; $(A\beta)=200$

Do you find the data consistent ?

SOLUTION. From the given data,

$N=600$; $(A)=250$; $(AB)=(A)-(A\beta)=250-200=50$

$(B)=(AB)+(\alpha B)=50+400=450$

Now, according to the conditions of consistence,

$(AB) \nless (A)+(B)-N$; or $(AB) \nless 250+450-600$; or $(AB) \nless 100$

But the value of AB is 50 which is less than 100. Hence the given data are inconsistent.

PROBLEM 10.5. *The following are the proportions per 1000 of boys observed by an investigator for certain classes of defects amongst a number of school children :*

A=development defects,
B=nerve signs,
C=mental dullness.

$N=1000$
$(A)=520$
$(B)=310$
$(C)=470$

$(AB)=80$
$(AC)=150$
$(BC)=85$
$(ABC)=25$

Show that the information as it stands must be incorrect.

SOLUTION. According to the conditions of consistence.

$(ABC) \ngtr (AB)+(AC)+(BC)-(A)-(B)-(C)+N.$

Substituting the values, we get,

$(ABC) \ngtr 80+150+85-520-310-470+1000$

or $ABC \ngtr 15$

But the given value of ABC is 25 which is more than 15. Hence the information must be incorrect.

PROBLEM 10.6. *A market investigator returns the following data of 10,000 people consulted :*

8110 liked chocolates ; 7520 liked toffee ;
4180 liked boiled sweets ; 5700 liked chocolates and toffee ;
3500 liked chocolates and boiled sweets, and 3480 liked toffee and boiled sweets, 2970 liked all three.

Show that this information as it stands must be incorrect.

SOLUTION. Denoting.

Persons who liked chocolates by A ; persons who liked toffee by B and persons who liked boiled sweets by C ;

the given data are :

$(A)=8110$
$(B)=7520$
$(C)=4180$
$(AB)=5700$

$(AC)=3500$
$(BC)=3480$
$(ABC)=2970$
$N=10000$

Now, according to the conditions of consistence,

$(ABC) \ngtr (AB)+(AC)+(BC)-(A)-(B)-(C)+N.$

Substituting the given values, we get,

$(ABC) \ngtr 5700+3500+3480-8110-7520-4180+10000$

or $(ABC) \ngtr 2870$

So ABC can be either equal to or less than 2870. But as per the data given it is 2970. Hence the data are incorrect.

PROBLEM 10.7. *From the office of a railway hospital, the following information is supplied to you :—*

55 percent of the patients are men, 60 per cent are "aged" (over 60), 75 per cent non-able-bodied, 37 per cent aged men, 43 per cent non-able-bodied men, and 20 per cent aged non-able-bodied men.

Do you think that the information as it stands is correct ?

SOLUTION. Denoting

The attribute 'men' by A ; the attribute 'aged' by B
and the attribute 'non-able-bodied' by C,
the given data are,

$N=100$; $(AB)=37$; $(A)=55$; $(AC)=43$
$(B)=60$; $(C)=75$; $(ABC)=20$

Now, according to the conditions of consistence,

$(ABC) \nless (AB)+(AC)-(A)$

Substituting the given values, we get,

$(ABC) \nless 37+43-55$; or $(ABC) \nless 25$

The values of (ABC) as per the given data is 20 which is less than 25. Hence the given information is incorrect.

PROBLEM 10.8. *1000 persons of London were asked by a B.B.C. investigator to give the nationality of the music they liked. He returns the following data :—*

570 liked English ; 650 liked French, and
480 liked German ; 440 liked English and French
360 liked French and German, and 240 liked English and German
125 liked all three.

Show that the information as it stands must be incorrect.

SOLUTION. Denoting

Persons who liked music of English nationality by A,
Persons who liked music of French nationality by B,
Persons who liked music of German nationality by C,
the given data :

$N=1000$; $(AB)=440$; $(A)=570$; $(AC)=240$
$(B)=650$; $(BC)=360$; $(C)=480$; $(ABC)=125$

Now according to the conditions of consistence,

$ABC \nless (AB)+(BC)-(B)$

Substituting the values of (AB), (BC) and B, we get,

$(ABC) \nless 440+360-650$; or $(ABC) \nless 150$

As per the data given the value of ABC is 125 which is $\nless$ 150. Hence the information is incorrect.

PROBLEM 10.9. *If a report gives the following frequencies as actually observed, show that there must be misprint or mistake of some sort, and that possibly the misprint consists in the dropping of a 1 before the 85 given as the frequency (BC) :—*

	$N=1{,}000$
$(A)=510$	$(AB)=189$
$(B)=490$	$(AC)=140$
$(C)=427$	$(BC)=85$

SOLUTION. According to the conditions of consistence.

$(AB)+(AC)+(BC) \nless (A)+(B)+(C)-N.$

or $(BC) \nless (A)+(B)+(C)-N-(AB)-(AC).$

Substituting the given values, we get,

$(BC) \nless 510+490+427-1000-189-140$ or $BC \nless 98$.

But the reported value of (BC) is 85 which is less than 98. Therefore 85 ca not be the correct value of (BC),

If we read 185 for 85, all the conditions of consistence are fulfilled.

PROBLEM 10.10. *Prove that the following data are inconsistent :—*

$(N) = 1{,}000 \qquad (A\beta) = 483$
$(A) = 525 \qquad (A\gamma) = 378$
$(B) = 312 \qquad (B\gamma) = 226$
$(C) = 470 \qquad (ABC) = 25$

SOLUTION. From the data given :

$(AB) = (A) - (A\beta) = 525 - 483 = 42$
$(AC) = (A) - (A\gamma) = 525 - 378 = 147$
$(BC) = (B) - (B\gamma) = 312 - 226 = 86$

According to conditions of consistence no class-frequency should be negati In the problem :

$\alpha\beta\gamma = N - A - B - C + AB + AC + BC - ABC$
$= 1000 - 525 - 312 - 470 + 42 + 147 + 86 - 25 = 1275 - 1332 = -57$

Thus, the data are inconsistent.

PROBLEM 10.11. *The following information is given to you :—*

100 children took three examinations. 40 passed the first, 39 passed the seco and 48 passed the third. 38 passed the first and the second. 35 passed the first and th third. 37 passed the second and the third.

Show that the information as it stands must be incorrect.

SOLUTION. Denoting.

Success in the first examination by A ; Success in the second examination by and Success in the third exam. by C ;

the given data are :

$N = 100; (AB) = 38 ; (A) = 40; (AC) = 35$
$(B) = 39; (BC) = 37 ; (C) = 48$

Now, according to the conditions of consistence,

$(AB) - (AC) + (BC) \nless (B)$ or $(BC) \nless (B) - (AB) + (AC)$

Substituting the given values, we get,

$(BC) \nless 39 - 38 + 35$ or $(BC) \ngtr 36$.

The given value of (BC) is 37 which is $\ngtr$ 36.

Hence the given information is incorrect.

PROBLEM 10.12. *The following summary appears in a report on a surv covering 1,000 fields. Scrutinize the numbers, and point out if there be any mistake, o misprint in them :—*

Manured	510
Irrigated fields	490
Fields growing improved varieties	427
Fields both irrigated and manured	189
Fields both manured and growing improved varieties ...	140
Fields, both irrigated and growing improved varieties ...	85

SOLUTION. Denoting

Manured fields by A ; irrigated fields by B and fields growing improved varieties by C, the given data are :

$N = 1000$ $(AB) = 189$
$(A) = 510$ $(AC) = 140$
$(B) = 490$ $(BC) = 85$
$(C) = 427$

According to the condition of consistence.

$(AB)+(AC)+(BC) \nless (A)+(B)+(C)-N.$

Substituting the above values for (A), (B), (C) and N, we get,

$(AB)+(AC)+(BC) \nless 510+490+427-1000$ or $(AB)+(AC)+(BC) \nless 427.$

Now, as per summary appearing in the report,

$(AB)+(AC)+(BC) = 189+140+85 = 414$, which is less than 427.

Hence there must be a misprint or mistake of some sort in the given figures.

PROBLEM 10.13. *The following is a summary of the statistical features of a census of ration cards :—*

Item No.	*Category*	*Total number of cards belonging to the category*
1	*The whole of the census*	*10,000*
2	*Permanent residents*	*5,100*
3	*Males*	*4,900*
4	*Consumers of rice*	*4,270*
5	*Permanent male residents*	*1,890*
6	*Consumers of rice among permanent residents*	*1,400*
7	*Males consuming rice*	*970*

Show that the entry against item No. 7 is inconsistent with the entries against all the previous items, namely, 1, 2, 3, 4, 5 and 6, taken together.

SOLUTION. Denoting

Permanent residents by A ; males by B, and consumers of rice by C, the given data are :—

$N = 10000$ $(AB) = 1890$
$(A) = 5100$ $(AC) = 1400$
$(B) = 4900$ $(BC) = 970$
$(C) = 4270$

Now, according to the conditions of consistence,

$(AB)+(AC)+(BC) \nless (A)+(B)+(C)-N$, or $(BC) \nless (A)+(B)+(C)-N-(AB)-(AC).$

Substituting the given values for (A), (B), (C), N, (AB) and (BC)

We get,

$$(BC) \nless 5100+4900+4270-10000-1890-1400$$
$$980.$$

The given value of (BC), is 970 which is less than 980. Hence, it can be said that the entry against item No. 7 is inconsistent with the entries against all the previous items, namely 1, 2, 3, 4, 5 and 6, taken together.

PROBLEM 10.14. *If in a village actually affected by anthrax, 70 per cent of the goats are attacked and 85 per cent have been inoculated with vaccine. What is the lowest percentage of the inoculated that must have been attacked ?*

SOLUTION. Denoting

the attribute of attack by A ;
and the attribute of inoculation by B ;

the given data are :
$(A)=70$; $(B)=85$; $N=100$

We have to find the lowest percentage of AB.

Now, according to the conditions of consistence,
$(AB)\nless 0$, and $(AB)\nless(A)+(B)-N$, *i.e.*, $70+85-100=55$.
Hence the value of AB cannot be less than 55.

Hence the lowest percentage of the inoculated that must have been attacked

$$\left(\frac{AB}{B}\times 100=\frac{55}{85}\times 100\right)=65 \text{ per cent.}$$

PROBLEM 10.15. *If, in the urban district of Bombay, 250 per thousand of women between 20 and 25 years of age were returned as "employed" at a census 850 per thousand as married or widowed, what is the lowest proportion per thousand the married or widowed that must have been employed ?*

SOLUTION. Let A be the attribute of employed women,
and B, that of the married or widowed women.

Then,
$(A)=250$; $(B)=850$; $N=1000$

Now,
in order that none of the ultimate frequency is negative,
$(AB)\nless 0$ and
$(AB)\nless$ $(A)+(B)-N$, *i.e.*, $250+850-1000$, or
$\nless 100$.

So (AB) cannot be less than 100.

Hence the lowest proportion per thousand of the married or widowed that m have been employed $\frac{AB}{B}\times 1000$ or $\frac{100}{850}\times 1000$ or 118.

PROBLEM 10.16. *The following are the proportions per 1000 of girls obser for certain classes of defects amongst a number of school-children,*

A=development defects, B=nerve signs, C=mental dullness.

$N=1000$ $(AB)=55$
$(A)=68$ $(BC)=36$
$(B)=85$
$(C)=69$

Show that some defectively developed girls are dull and state how many at lea must be so.

SOLUTION. In the above question, the least possible value of (AC) has to found out,

Now, according to the conditions of consistence,
$(AB)-(AC)+(BC)\ngtr(B)$ or $(AB)+(BC)\nless(B)+(AC)$
or $(B)+(AC)\ngtr(AB)+(BC)$ or $(AC)\nless(AB)+(BC)-B$,
or $(AC)\nless 55+36-85$ or 6.

Hence, there must be at least 6 defectively developed girls who are dull.

PROBLEM. 10.17. *The following are the proportions per 5,000 of worker observed for certain classes of defects amongst a number of factory workers.*

A=development defects ; B=nerve signs ; D=mental dullness.
$N=5000$; $(D)=400$; $(A)=440$; $(AB)=170$; $(B)=545$; $(BD)=228$.

Show that some dull workers do not exhibit development defects and state ho many at least do not do so.

SOLUTION. In the above question, we have to find out the least possible ue of (αD).

So, we should find out the greatest of possible value of AD

According to the conditions of consistence,

(1) AB+AD—(BD)$\not<$(A) or (AD)$\not>$(A)+(BD)—(AB)
Substituting the given values of (A), (BD) and (AB) we get,
(AD) $\not>$440+228—170 or (AD)$\not>$498.

(2) Similarly,

—(AB)+(AD)+(BD)$\not<$D or (AD)$\not>$D+(AB)—(BD)
Substituting the given values of (D), (AB) and (BD), we get,
(AD)$\not<$400+170—228 or (AD)$\not>$342.

So, the greatest possible value of (AD) is 342.
nce,

(αD)=(D)—(AD) or (αD)=400—342=58.

Thus the least number of dull workers who do not exhibit development ects is 58.

PROBLEM 10.18. *Given that 50 per cent of the inmates of college hostel are ls, 60 per cent are between teens, 80 per cent intelligent, 35 per cent girls between ns, 45 per cent intelligent girls, and 42 per cent between teens and intelligent. d the greatest and least possible proportions of intelligent girls between teens.*

SOLUTION. Denoting the attributes 'girls', 'between teens' and 'intelligence' A, B and C, respectively, and letting N=100, the given data are :

(A)=50; (AB)=35 : (B)=60; (AC)=45 ; (C)=80; (BC)=42.

The greatest and the least proportions of (ABC) have to be found out. cording to one set of conditions of consistence :

(ABC)$\not<$0

(ABC)$\not<$(AB)+(AC)—(A), or, 35+45—50, or 30.

(ABC)$\not<$(AB)+(BC)—(B), or, 35+42—60, or, 17.

(ABC)$\not<$(AC)+(BC)—(C), or, 45+42—80, or 7.

So, the least possible value of ABC is 30. According to another set of condi- ns of consistence,

(ABC)$\not>$(AB) *i.e.* 35

(ABC)$\not<$(AC) *i.e.* 45

(ABC)$\not>$(BC) *i.e.* 42

(ABC)$\not>$(AB)+(AC)+(BC)—(A)—(B)—(C)+N.
$\not>$35+45+42—50—60—80+100.
$\not>$32.

So the greatest possible value of ABC is 32.

PROBLEM 10.19. *Among the adult population of a certain town 50 per cent of e population are male, 60 per cent are wage earners and 50 per cent are 40 years f age or over, 10 per cent of the males are not wage earners and 40 per cent of the ales are under 42. Can we infer anything about what percentage of the population f 40 years or over are wage earners ?*

SOLUTION. Denoting the attributes of male, wage earner and 40 years old more by A, B and C respectively, and letting N=100, the given data are :

(A)=50 ; (B)=60 ; (C)=50 ; (Aβ)=5 ; (Aγ)=20
From the above
N=100; (AB)=(A)—(Aβ)=50—5=45.
(A)=50; (AC)=(A)—(Aγ)=50—20=30
(B)=60; (C)=50

We are required to find out the limits of (BC). According to the conditi of consistence.

Firstly,

$(AB)+(AC)+(BC) \nless (A)+(B)+(C)-N,$

Or $\quad (BC) \quad \nless (A)+(B)+(C)-N-(AB)-(AC)$

Or $\quad (BC) \quad \nless 50+60+50-100-45-30$

Or $\quad (BC) \quad \nless -15$ (1)

Secondly,

$(AB)+(AC)-(BC) \ngtr (A)$; Or $(AB)+(AC) \ngtr (A)+(BC)$

Or $(BC) \nless (AB)+(AC)-(A)$; Or $(BC) \nless 45+30-50$

Or $\quad (BC) \nless 25$ ... ... (2)

Thirdly,

$(AB)-(AC)+(BC) \ngtr (B)$ Or $(BC) \ngeq (B)+(AC)-(AB)$; (3)

Or $\quad (BC) \ngtr 60+30-45$; Or $(BC) \ngtr 45$

Fourthly,

$-(AB)+(AC)+(BC) \ngtr (C)$; Or $(BC) \ngtr (C)+(AB)-(AC)$ (4)

Or $\quad (BC) \ngtr 50+45-30$; Or $\quad (BC) \ngtr 65$

From generalisations (1) and (2), the lowest possible value of (BC) is 25 a from generalisations (3) and (4), the greatest possible value of (BC) is 65.

Hence, the percentage of the population 40 years old or more who are w earners $\left(\text{that is, } \frac{BC}{C} \times 100\right)$ lies between 50 and 90

PROBLEM 10.20. *50% of the imports of barley into a country come from commonwealth countries ; 80% of the total imports go to brewing ; 75% of imports are grown in the Northern·hemisphere ; 80% of the Northern grown barley to brewing ; 100% of the foreign Southern grown barley goes to stock feeding. that the foreign Northern grown barley which goes to brewing cannot be less than nor more than 60% of the total imports. (It is assumed that brewing and stock feedi are the only two uses to which imported barley is put).*

SOLUTION. Denoting

Imports from Commonwealth countries by A,
Going to Brewing by B,
Grown in Northern hemisphere by C, we get,

$(N)=100$; $(A)=50$; $(B)=80$; $(C)=75$

$$(BC)=\left(\frac{80}{100} \times 75\right)=60$$

As 100% of foreign Southern grown barley goes to stock feeding

$\therefore (\alpha\gamma)=(\alpha\beta\gamma)$

We have to prove that $(\alpha BC) \quad 30 \ngtr 50$

$(ABC)=(BC)-(\alpha BC)-60-(\alpha BC)$

$(\alpha\beta\gamma)=(\alpha\gamma)$ or

$N-(A)-(B)-(C)+(AB)+(AC)+(BC)-(ABC) =N-(A)-(C)+(AC)$

or $\quad (AB)+(BC) \quad =(ABC)+(B)$

or $\quad (AB) \quad =(ABC)+(B)-(BC)=60-(\alpha BC)+80-60$

$\qquad =80-(\alpha BC)$

$(A\beta)=(A)-(AB) =50-\{80-(\alpha BC)\}$

$\qquad =50-80+(\alpha BC)=(\alpha BC)-30$

Since $(A\beta)$ cannot be negative it is clear that (αBC) cannot be less than 30.

$(\alpha\beta)=N-(A)-(B)+(AB)$

$=100-50-80+80-(\alpha BC)$
$=50-(\alpha BC)$

Since $(\alpha\beta)$ cannot be negative it is clear that (αBC) cannot be more than 50.

PROBLEM 10.21. *A coin is tossed three times and the results heads and tails noted. The process is continued until there are 100 sets of threes. In 69 cases heads fell first, in 49 cases heads fell second and in 53 cases heads fell third. In 33 cases heads fell both first and second and in 21 cases heads fell both second and third. Show that there must have been at least 5 occasions on which heads fell three times, and that there could not have been more than 15 occasions on which tails fell three times, though there need not have been any.*

SOLUTION. Denoting

Heads falling first by A
Heads falling second by B
Heads falling third by C

We get,

$(A)=69$; $(B)=49$; $(C)=53$; $(AB)=33$; $(BC)=21$ and $N=100$

We have to prove that,

$(ABC) \nless 5$ and $(\alpha\beta\gamma) \ngtr 15$ though it may have a value zero

(i) Now $(ABC) \nless (AB)+(BC)-(B)$
$\nless 33+21-49$
$\nless 5$ (Proved)

(*ii*) $(\alpha\beta\gamma)=N-(A)-(B)-(C)+(AB)+(BC)+(AC)-(ABC)$
$=N-(A)-(B)-(C)+(AB)+(BC)+(A\beta C)$
$=100-69-49-53+33+21+(A\beta C)$
$=(A\beta C)-17$
$(\beta C)=(C)-(BC)=53-21=32$

$(A\beta C) \ngtr (\beta C) \ngtr 32$

As such, $(\alpha\beta\gamma) \ngtr 32-17$ or 15 (Proved)

(*iii*) $(AC) \nless (A)+(C)-N$
$\nless 69+53-100$
$\nless 22$

Since $(\alpha C)=(C)-(AC)$ and since (AC) cannot be less than 22 it follows that $(\alpha C) \ngtr 53-22$ or 31.

Supposing the actual value of (αC) happens to be 31 then,
$(\alpha\gamma)$ could be equal to $(\alpha)-(\alpha C)$ or $31-31$ or O

If $(\alpha\gamma)=O$ then $(\alpha\beta\gamma)$ would also be O. As such $(\alpha\beta\gamma)$ may have a value which is zero.

REVISIONAL PROBLEMS—10

PROBLEM 10.22. From the following ultimate class frequencies, find the frequencies of the positive and negative classes and the total number of observations :

$(AB)=100$; $(\alpha\beta)=40$; $(\alpha B)=80$; $(A\beta)=50$.
[*Ans* : $(A)=150$: $(B)=180$; $(\alpha)=120$; $(\beta)=90$; $N=270$].

PROBLEM 10.23. Out of 70,000 literates in a particular district in India the number of criminals was 500. Out of 9,30,000 illiterates in the same district, number of criminals was 15,000. On the basis of these figures do you find any association between illiteracy and criminality ?

[*Ans* : $\frac{(AB)}{(A)} \times 100 = 1.6$; $\frac{(\alpha\beta)}{(\alpha)} \times 100 = .7\%$; *Association is positive*].

PROBLEM 10.24. Define the coefficient of association of two attribute Obtain such a measure between unemployment and educational attainments fro the following results of urban survey :

	Employed	Unemployed
Illiterate or below matric	5997	532
Matric and above	572	96

[*Ans* : $Q = +0.399$

PROBLEM 10.25. In a report on consumers preference it was given that o of 500 persons surveyed 410 preferred variety A, 380 preferred variety B and 2 persons liked both. Are the data consistent ?

[*Ans* : $(\alpha\beta) = -20$, *No*]

PROBLEM 10.26. Calculate $(\alpha\beta\gamma)$, $(A\beta C)$, $(\alpha\gamma)$ and $(A\beta)$ from the da given below :

$N=100$, $(A)=60$, $(B)=70$, $(C)=75$, $(AB)=30$, $(BC)=50$
$(AC)=45$, $(ABC)=20$

[*Ans* : *3, 25, 10 and 30*]

PROBLEM 10.27. 100 children took three examinations. 40 passed the first, 39 passed the second and 48 passed the third, 10 passed all the three, 21 failed all the three, 9 passed 1st two and failed in third. Find how many children passed at least two examinations.

[*Ans* : *38*].

PROBLEM 10.28. A survey of 1000 companies give the following data.

(i)	Companies with a capital of more than Rs. 10 lakhs	=510
(ii)	Companies making profit	=490
(iii)	Companies under Managing Agents	=427
(iv)	Companies with a Capital of more than Rs. 10 lakhs and making profit	=189
(v)	Cos. with a Capital of more than Rs. 10 lakhs and under Managing Agents	=140
(vi)	Cos. making profits and under Managing Agent	= 85

Show that the information as it stands must be incorrect ?

[*Ans* : *BC cannot be less than 98 and hence information is incorrect*].

PROBLEM 10.29. At an examination at which 600 candidates appeared, boys outnumbered girls by 16% of all candidates. Number of passed candidates exceeded the number of failed candidates by 310. Boys failing in the examination numbered 88. Construct the two fold association table and calculate the coefficient of association between male sex and success in the examination.

[*Ans* : $Q = -.07$]

PROBLEM 10.30. Calculate the coefficient of contingency from the following data :

Social Status	Dull	Intelligent	Brilliant
Lower Middle	22	35	23
Middle	38	70	32
Upper Middle	60	20	20

[*Ans* : *0.315*]

PROBLEM 10.31. Do you find any inconsistency in the data given below :
$N=2000$; $(A)=1000$, $(B)=1000$, $(C)=900$, $(AB)=400$, $(BC)=500$
$(AC)=300$, $(ABC)=240$

[*Ans* : *Yes*, $(\alpha\beta\gamma = -40)$]

Probability

11

Till now we have solved questions relating to such data which are obtained from sources over which we have no control, *e. g.*, fluctuations in prices, distribution of income, etc. In this chapter we shall solve questions relating to artificial data which have been collected from such sources over which we have more or less full control. Such data are analyzed to study the statistical laws from which these figures emerge. This discussion relates to probability.

If an event can happen in a ways and fail in b ways, and each of these ways is equally likely, the probability of its happening is equal to $\frac{a}{a+b}$. If the events are mutually exclusive (*i.e.* if one event happens the other cannot happen) then the probability that some one of these events will occur is equal to the total of the probabilities of occurrence of various events. That is called addition of probabilities. If on the other hand two or more events can occur simultaneously and if the events are independent of each other then the probability that all of them will happen is equal to the product of the probabilities of the various events. This is called multiplication of probabilities.

In this chapter various types of questions dealing with simple probability, rule of addition, rule of multiplication, mathematical expectation and inverse probability have been solved.

PROBLEM 11.1. *An urn contains 2 black balls and 3 white balls. Find out the probability that a blind-folded person in one draw shall obtain a black ball.*

SOLUTION. If an event can happen in a ways and fail in b ways, and each of these ways is equally, likely, the probability of its happening or $p=a/(a+b)$.

In the given problem, $a=2$ and $b=3$; since the blind-folded person can obtain a black ball in 2 ways and can fail to obtain a black ball in 3 ways.

Therefore, the probability of his obtaining a black ball is

$$\frac{2}{2+3}=\frac{2}{5}$$

PROBLEM 11.2. *A group of scientific men reported 11705 sons and 11527 daughters. If this is a fair sample from the general population, what is the probability that a child to be born will be a boy ?*

SOLUTION. The event can happen in 11705 ways and can fail to happen in 11527 ways.

Therefore the probability of the event that a child to be born will be a boy is $\frac{11705}{11705+11527}$, or $\frac{11705}{23232}$

PROBLEM 11.3. *What is the probability that a vowel selected at random in an English book is an e ?*

SOLUTION. If e is the total number of cases, each being equally likely to occur, and of these a are favourable to the event then the probability that the event will happen is $\frac{a}{e}$

In the given problem, the total number of cases is 5, since the selected vowel may be either a, or e, or i, or o, or u.

And the case favourable to the event is 1 ;

Therefore, the probability that a vowel selected at random in an English book is an *e* is $\frac{1}{5}$.

PROBLEM 11.4. *The American Experience Table states that of a population of 100,000 alive at age ten, 49,341 are alive at age 65, Calculate the probability of surviving for this period for a person who is 10 years of age at present.*

SOLUTION. Cases favourable to the event, that is, survival for the period are 49,341.

and, the total number of cases is 1,00,000.

Therefore, the probability of surviving for the said

$$\text{person at the age of } 10 = \frac{49341}{100000}$$

PROBLEM 11.5. *What is the probability that a letter selected at random in an English book is a vowel ?*

SOLUTION. Take a random page in a random English book and count the number of letters and the number of vowels.

$$\text{The probability is} = \frac{\text{No. of vowels}}{\text{No. of letters}}$$

PROBLEM 11.6. *According to the Life Tables, out of one lakh of persons living at age 10, 82,134 survive to age 40, of whom 837 die in a year. What is the probability for a man of 40 of surviving one more year ?*

SOLUTION. At the age of 40, the total number of survivors is 82,134, of these 837 die in a year.

That is, (82134—837)=81,297 persons aged 40 survive one year more.

Therefore, the probability that a person aged 40 would survive a year is

$$= \frac{81,297}{82,134}$$

PROBLEM 11.7. *In 128 litters, each of five puppies, the number of males were distributed as follows :*

No. of males per litter	*0*	*1*	*2*	*3*	*4*	*5*
No. of litters	*3*	*22*	*44*	*38*	*16*	*5*

Calculate the probability of a male birth.

SOLUTION. Total number of puppies$=128 \times 5=640$.

$$\text{Male puppies in all the litters} = (3\times0)+(22\times1)+(44\times2)+(38\times3)+(16\times4)+(5\times5)$$

$$=0+22+88+144+64+25=313.$$

$$\text{Therefore, the probability of male birth is} = \frac{313}{640}$$

PROBLEM 11.8. *Eight balls numbered from 1 to 8 are placed in a bag and two are drawn at random. What is the probability that they are numbered 1 and 2 ?*

SOLUTION. Total number of ways in which two balls can be drawn out of 8 balls is ${}^8C_2$, or $\frac{8\,!}{(8-2)\,!\times 2\,!}$

$$\text{or } \frac{8\times7\times6\times5\times4\times3\times2\times1}{6\times5\times4\times3\times2\times1\times2\times1}=28$$

The possible arrangement of two balls numbered 1 and 2 is 1.

Therefore, the probability that 1 and 2 numbered balls are drawn in one draw is $\frac{1}{28}$.

PROBLEM 11.9. *A bag contains 6 white, 7 red and 5 black balls ; find the chance that three balls drawn at random are all white.*

SOLUTION. Three balls can be drawn out of $(6+7+5)=18$ balls.

in ${}^{18}C_3$ or $\frac{18\,!}{(18-3)\,!\times 3!}$ or $\frac{18\times 17\times 16}{3\times 2\times 1}=816$ ways

And, 3 balls can be drawn out of 6 white balls.

in ${}^{6}C_3$ or $\frac{6}{(6-3)\,!\,\times 3!}=\frac{6\times 5\times 4}{3\times 2\times 1}=20$ ways.

Therefore, the chance that three balls drawn at random

are all white is $=\frac{20}{816}=\frac{5}{204}$.

PROBLEM 11.10. *Find out the probability that a man asked to form a two-digit number out of 2, 3, 5, 7, 9 would form 79, when,*

(*a*) *repetitions are not allowed.* (*b*) *repetitions are allowed.*

SOLUTION. (*a*) *When repetitions are not allowed.*

Out of the given five numbers, two digit numbers can be formed in ${}^5p_2$, or

$\frac{5\;!}{(5-2)\;!}$ or $5\times 4=20$ ways.

Of these, 79 is one possibility,

Therefore the probability that a man asked to form a two-digit number out of 2, 3, 5, 7 and 9 would form 79 is $=\frac{1}{20}$

(*b*) *When repetitions are allowed.*

Out of the given five numbers, two digit numbers can be formed in 5×5, o 25 ways.

Of which 79 is one possibility.

Therefore, the required probability in this case is $=\frac{1}{25}$.

PROBLEM 11.11. *What is the chance of throwing a number greater than with an ordinary die whose faces are numbered from 1 to 6 ?*

SOLUTION. There are 6 possible ways in which the die can fall, and o these two are favourable to the event required ;

Therefore the required chance$=\frac{2}{6}=\frac{1}{3}$

PROBLEM 11.12. *In a single throw with two dice find the chance of thro wing (1) eight (2) eleven.*

SOLUTION. (1) A throw amounting to 8 must be made up of 2, 6 ; or 5 ; each of which arrangement can occur in 2 ways ; or 4, 4 which can occur in way only.

Therefore the number of favourable cases is $2+2+1$, or 5.

And the total number of cases is 6^2, or 36,

therefore the required chance$=\frac{5}{36}$

(2) A throw amounting to 11 must be made up of 6, 5 which can occur in ways.

And, the total number of cases is 6^2, or 36,

therefore, the required chance $=\frac{2}{36}=\frac{1}{18}$

PROBLEM 11.13. *An ordinary six-sided dice is thrown 4 times ; what a the probabilities of obtaining 4, 3, 2, 1, 0 aces ?*

SOLUTION. The probabilities of obtaining 4, 3, 2, 1, 0 aces would be the values of the various terms of the expansion of $(\frac{1}{6}+\frac{5}{6})^4$ Thus,

The probability of 4 aces = 1/1296
The probability of 3 aces = 20/1296
The probability of 2 aces = 150/1296
The probability of 1 ace = 500/1296
The probability of 0 ace = 625/1296

PROBLEM 11.14. *Twelve ordinary dice are thrown once. What is the probability of obtaining exactly 4 aces ?*

SOLUTION. The probability of 4 aces.

$$= {}^{12}C_4\, p^4 q^8 \text{ where } p = \frac{1}{6} \text{ and } q = \frac{5}{6}$$

$$= \frac{12!}{8!4!}\left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^8 = \frac{1}{11}$$

PROBLEM 11.15. *An urn contains 7 green and 5 black balls. Four are drawn at random. A green ball is a success. Find the probabilities of 4, 3, 2, 1, 0 green balls.*

SOLUTION. Probabilities of 4 green balls $\quad {}^7C_4 \div {}^{12}C_4 = 7/99$

" " 3 " " $\quad {}^7C_3 \times {}^5C_1 \div {}^{12}C_4 = 35/99$

" " 2 " " $\quad {}^7C_2 \times {}^5C_2 \div {}^{12}C_4 = 42/99$

" " 1 " " $\quad {}^7C_1 \times {}^5C_3 \div {}^{12}C_4 = 14/99$

" " 0 " " $\quad {}^5C_4 \div {}^{12}C_4 = 1/99$

99/99

PROBLEM 11.16. *Find the chance of throwing at least an ace in a single throw with two dice.*

SOLUTION. The possible number of cases is 6×6, or 36.

An ace on one die may be associated with any of the 6 numbers of the other die, and the remaining 5 numbers on the first die may each be associated with the ace on the second die; thus the number of favourable cases is 11.

Therefore the required chance is $= \frac{11}{30}$

PROBLEM 11.17. *Compare the chances of throwing 4 with one die, 8 with two dice, and 12 with three dice.*

SOLUTION. (*1*) *Chances of throwing 4 with one die.*

There are 6 possible ways in which the die can fall, and of these one is favourable to the event required.

therefore the required chance is $= \frac{1}{6}$.

(*2*) *Chance of throwing 8 with two dice.*

To make a throw of 8, the two must fall either 2, 6, or 3, 5 or 4, 4, of which the first two arrangements can occur in two ways and the third arrangement in one way only.

Therefore, the number of favourable cases is 2+2+1, or 5.

And the total number of cases is 6×6, or 36;

therefore the required chance is $\frac{5}{36}$

(*3*) *Chances of throwing 12 with three dice.*

To make a throw of 12, the three dice must fall either 6, 1, 5, or 6, 2, 4; or 6, 3, 3, or 5, 2, 5; or 5, 3 ; 4; or 4, 4, 4; the first two of which arrangements can occur in 3 ! ways; the second two in 3 ways each, the fifth in 3 ! ways and the last in one way only.

Therefore, the number of favourable cases is

$$6+6+3+3+6+1, \text{ or } 25,$$

And the total number of cases is 6^3, or 216

$$\text{therefore the required chance} = \frac{25}{216}$$

Therefore, the chances of throwing 4 with one die, 8 with two dice, and 12 with three dice are $\frac{1}{6} : \frac{5}{36} : \frac{25}{216}$; or 36 : 30 : 25.

PROBLEM 11.18. *A and B throw with three dice; if A throws 14. find B's chance of throwing a higher number.*

SOLUTION. To throw higher number than A, B must throw either 15, or 16, or 17 or 18.

A throw amounting to 18 must be made up of 6, 6, 6, and thus can occur in one way; 17 can be made up of 6, 6, 5 which can occur in 3 ways: 16 may be made up of 6, 6, 4 and 6, 5, 5, each of which arrangements can occur in 3 ways; 15 can be made up of 6, 4, 5; or 6, 3, 6; or 5, 5, 5 which can occur in 3 !, 3 and 1 way respectively.

Therefore the number of favourable cases is

$$1+3+3+3+6+3+1, \text{ or } 20$$

And the total number of cases is 6^3, or 216;

$$\text{therefore the required chance } \frac{20}{216} = \frac{5}{54}$$

PROBLEM 11.19. *From a pack of 52 cards two are drawn at random ; find the chance that one is a king and the other a queen.*

SOLUTION. From a pack of 52 cards, a draw of 2 can be made in ${}^{52}C_2$ ways. A king can be drawn in ${}^4C_1$ ways and similarly a queen in ${}^4C_1$ ways. Since each of the former can be associated with each of the latter, a king and a queen can be drawn in ${}^4C_1 \times {}^4C_1$ ways,

$$\text{Therefore, the required chance} = \frac{{}^4C_1 \times {}^4C_1}{{}^{52}C_2} = \frac{4 \times 4 \times 2}{52 \times 51} = \frac{8}{663}$$

PROBLEM 11.20. *If from a pack four cards are drawn, find the chance that they will be four honours of the same suit.*

SOLUTION. Four honours of the same suit can be drawn in 4 ways, since, there are 4 suits each containing four honours and from the pack, a draw of 4 cards can be made in,

$${}^{52}C_4, \text{ or } = \frac{52 \times 51 \times 50 \times 49}{4 \times 3 \times 2 \times 1} = 270725 \text{ ways.}$$

$$\text{Therefore, the required chance} = \frac{4}{270725}$$

PROBLEM 11.21. *In a hand at whist what is the chance that the 4 kings are held by a specified player?*

SOLUTION. In a hand at whist, the specified player must have 13 cards. If he possesses 4 kings, he should also have 9 cards of other type or types.

Thus, the possible arrangements of this sort are ${}^4C_4 \times {}^{48}C_9$. And from a pack of 52 cards, a distribution of 13 can be made in ${}^{52}C_{13}$ ways,

$$\text{Therefore the required chance} = \frac{{}^4C_4 \times {}^{48}C_9}{{}^{52}C_{13}} = \frac{11}{4165}$$

PROBLEM 11.22. *What is the probability that each of the four players of cards holds 13 cards of a specified same suit.*

SOLUTION. 52 cards can be distributed to 4 people in

$$= \frac{52\,!}{13\,!\,13\,!\,13\,!\,13\,!} \text{ ways}$$

of these in only one way in which the four people can hold 13 cards of a specified suit.

$$\text{Therefore, the required chance} = 1 \bigg/ \frac{52\,!}{13\,!\,13\,!\,13\,!\,13\,!}$$

$$= \frac{13\,!\,13\,!\,13\,!\,13\,!}{52\,!}$$

PROBLEM 11.23. *In shuffling a pack of cards, four are accidentally dropped; find the chance that the missing cards should be one from each suit.*

SOLUTION. There are ${}^{52}C_4$ possible ways in which four cards can slip while shuffling a pack of cards. The possible number of cases in which the four cards can be one from each suit is ${}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1$.

$$\text{Therefore the required chance} = \frac{{}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1}{{}^{52}C_4} = \frac{2197}{20825}$$

PROBLEM 11.24. *Four cards are drawn without replacement. What is the probability that they are all aces ?*

SOLUTION. The required probability of four aces would be :—

$$\frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{1}{270725}$$

PROBLEM 11.25. *Four cards are drawn without replacement. What is the chance that they all have the same value ?*

SOLUTION. The probability would be,

$$1 \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{1}{20825}$$

The first card may be any one of the 52. The second card should be any one of the three cards (of the same value as the first one) out of the remaining 51, and so on.

PROBLEM 11.26. *Four cards are drawn without replacement. What is the chance that they are all of different suits ?*

SOLUTION. The required probability would be,

$$1 \times \frac{39}{51} \times \frac{26}{50} \times \frac{13}{49} = \frac{2197}{20825}$$

The first card may be any one of the 52. The second card should be any one of the 39 cards (excluding the 13 cards of the suit to which the first card belonged) out of the remaining 51 and so on.

PROBLEM 11.27. *If* n *people are seated at a round table, what is the probability that two named individuals will be neighbours?*

SOLUTION. If one of the named individuals is seated, the other one can sit in $(n-1)$ places. Out of these, two seats would make him a neighbour of the first named individual.

Hence the required possibility $= 2/n(-1)$.

PROBLEM 11.28. *There are three works, one consisting of 4 volumes, one of 5, and the other of 2 volumes. They are placed on a shelf at random; find what is the chance that volumes of the same work are all together.*

SOLUTION. The number of ways in which the different volumes can be arranged.

$$\frac{(4+5+2)!}{4!\times5!\times2!}=\frac{11\times10\times9\times8\times7\times6}{4\times3\times2\times1\times2\times1}=6930$$

And the number of ways in which the volumes of the same work will be together is 3 ! , or 6.

Therefore the required chance$=\frac{6}{6930}=\frac{1}{1155}$

PROBLEM 11.29 *In a bag there are five white and three black balls. What is the probability that if they are drawn out one at a time the first will be white, the second black, the third white, the fourth black and so on ?*

SOLUTION. Eight balls, 5 being white and 3 black can be arranged in $\frac{8\,!}{5\,!\,3\,!}$ or 56 ways.

Of these the given arrangement that the first will be white; the second black, the third white, the fourth black and so on is one of the possibilities.

Hence the desired probability is$=\frac{1}{56}$.

PROBLEM 11.30. *A bag contains 4 white, 5 red, and 6 green balls. Three balls are drawn at random. What is the probability that a white, a red and a green ball are drawn ?*

SOLUTION

The number of ways in which 3 balls may be drawn is $^{15}C_3$

The number of ways in which 1 white ball may be drawn is $^{4}C_1$

The number of ways in which 1 red ball may be drawn is $^{5}C_1$

The number of ways in which 1 green ball may be drawn is $^{6}C_1$.

Thus, the number of ways in which 1 white, 1 red and 1 green ball may be drawn is $^{4}C_1\times{}^{5}C_1\times{}^{6}C_1$.

Hence the desired probability$=\frac{^{4}C_1\times{}^{5}C_1\times{}^{6}C_1}{^{15}C_3}$

$$=\frac{4\times5\times6\times3\times2}{15\times14\times13}=\frac{24}{91}$$

PROBLEM 11.31. *In tossing a coin find the chance of throwing head and tail alternately in 3 successive trials.*

SOLUTION. The first throw must give either head, or tail ; the chance that the second gives the opposite to the first is $\frac{1}{2}$, and the chance that the third throw is the same as the first is $\frac{1}{2}$.

Therefore the chance of the compound event$=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$.

PROBLEM 11.32. *Find the chance of drawing a king, a queen and a knave in that order from a pack of cards in three consecutive draws, the cards drawn not being replaced.*

SOLUTION.

The probability of drawing a king$=\frac{4}{52}$

the ,, ,, ,, a queen then$=\frac{4}{51}$

and ,, ,, ,, ,, a knave$=\frac{4}{50}$

Since they are dependent events, the required

chance of the compound event $= \frac{4}{52} \times \frac{4}{51} \times \frac{4}{50} = \frac{8}{16575}$

PROBLEM 11.33. *Three cards are drawn ; what is the chance that they are a king, queen and a knave in any order.*

SOLUTION. The probability of drawing any one of the 12 court cards.

(king, queen or knave) $= \frac{12}{52}$

The probability of drawing any one of the 8 cards.

(leaving the one set which has been drawn) $= \frac{8}{51}$

The probability of drawing any one of the 4 cards,

(leaving 2 sets which have been drawn) $= \frac{4}{50}$

The combined probability $= \frac{12}{52} \times \frac{8}{51} \times \frac{4}{50} = \frac{15}{5525}$

PROBLEM 11.34. *A card is drawn from a pack. If it is a king, queen or knave it is replaced ; otherwise not. A second card is then drawn.*

What is a man's expectation if he is to get one rupee for drawing a king, queen or knave at the first draw and rupees two for drawing a king, queen or knave at the second draw ?

SOLUTION. (*i*) First draw : Chance of a court card (King, queen or knave) $= \frac{12}{52}$

$\therefore$ The expectation $= \frac{12}{52} \times 1 = \frac{3}{13}$ rupees,

(*ii*) Second draw :

(*a*) If the result of the first draw were a court card (so that the card was replaced).

The probability of a court card again $= \frac{12}{52} \times \frac{12}{52}$

$\therefore$ The expectation $= \frac{12}{52} \times \frac{12}{52} \times 2 = \frac{18}{169}$ rupees

(*b*) If the result of the first draw were not a court card (so that the card was not replaced).

The probability of a court card $= \frac{40}{52} \times \frac{12}{51}$

$\therefore$ The expectation $= \frac{40}{52} \times \frac{12}{51} \times 2 = \frac{80}{221}$ rupees

The total expectation $= \frac{3}{13} + \frac{18}{169} + \frac{80}{221}$ rupees.

$= .70$ rupee approx.

PROBLEM 11.35. *A bag contains 6 white and 9 black balls. Two drawings f 4 balls are made such that (a) the balls are replaced before the second trial (b) he balls are not replaced before the second trial. Find the probability that the first rawing will give 4 white and the second 4 black balls in each case.*

SOLUTION. (*a*) *When the balls are replaced before the second trial.*

The number of ways in which 4 balls may be drawn is $^{15}C_4$;

The number of ways in which 4 white balls may be drawn is $^{6}C_4$;

The number of ways in which 4 black balls may be drawn is $^{9}C_4$;

Therefore the probability of drawing 4 white at the first trial $= {}^{6}C_4/{}^{15}C_4$

$$=\frac{6\times5}{2}\times\frac{4\times3\times2\times1}{15\times14\times13\times12}=\frac{1}{91}\ ;$$

and the chance of 4 black at the second trial $= {}^{9}C_4/{}^{15}C_4$

$$=\frac{9\times8\times7\times6}{4!}\times\frac{4\ !}{15\times14\times13\times12}=\frac{6}{65}\ ;$$

therefore the chance of the compound event $=\frac{1}{91}\times\frac{6}{65}=\frac{6}{5915}$

(*b*) *When the balls are not replaced.*

At the first trial, 4 balls may be drawn in $^{15}C_4$ ways and 4 white balls may be drawn in $^{6}C_4$ ways ; therefore the chance of 4 white balls at first is $^{6}C_4/{}^{15}C_4$

$$=\frac{1}{91}$$

When 4 white balls have been drawn and removed, the bag contains 2 white and 9 black balls.

Therefore at the second trial,

4 balls may be drawn in $^{11}C4$ ways ; and 4 black balls may be drawn in $^{9}C_4$ ways. Therefore the chance of 4 black at the second trial.

$$={}^{9}C_4/{}^{11}C_4=\frac{9\times8\times7\times6}{4\ !}\times\frac{4!}{11\times10\times9\times8}=\frac{21}{55}\ ;$$

therefore the chance of the compound event,

$$=\frac{1}{91}\times\frac{21}{55}=\frac{3}{715}$$

PROBLEM 11.36. *Lloyd, the captain of the West Indies cricket team, is reported to have observed the rule of calling 'heads' every time the toss was made during the five matches of the last test-series with the Indian team. What is the probability of his winning the toss in all the five matches?*

How will the probability be affected if he had made a rule of tossing a coin privately to decide whether to call 'heads' or 'tails' on each occasion.

SOLUTION. The chance of winning the toss by Lloyd, the captain of the West Indies cricket team, in one test is $\frac{1}{2}$,

Since his winning the toss in any test is not affected by the results of other tosses in other tests, the events under consideration are independent.

Hence the probability of his winning the toss in all the five

matches $=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{32}$

The calculated probability would not be affected even if he had made a rule of tossing a coin privately to decide whether to call . 'head' or 'tail' on each occasion, because the events under consideration are independent of each other.

PROBLEM 11.37. *A bag contains 19 counters marked with the numbers 1 to 19. A counter is drawn and replaced ; a second drawing is then made, what is the chance that the first number drawn is even and the second odd ?*

SOLUTION. Out of 19 counters marked with the numbers 1 to 19, there are 9 counters marked with even numbers and 10 counters marked with odd numbers, since the numbers 2, 4, 6, 8, 10, 14, 16, 18 are even and the rest odds.

Thus, the chance of first drawing an even-numbered counter $=\frac{9}{19}$

and the chance of then drawing an odd-number counter $=\frac{10}{19}$

Therefore the chance of the compound event.

$$=\frac{9}{19}\times\frac{10}{19}=\frac{90}{361}$$

PROBLEM 11.38. *A lady declares that, by tasting a cup of tea made with milk, she can discriminate whether milk or tea infusion was first made into the cup. It is proposed to test this claim by means of an experiment with 12 cups of tea, 6 made in one way and 6 in the other way, and presenting them, in random order to her.*

Calculate the probability that on the null hypothesis the lady would judge correctly all the cups it being known to her that 6 are of each kind.

If, however, the 12 cups were presented to the lady in 6 pairs, each pair to consist of either kind, and the presentation be again in random order, how will the probability of correctly judging with every cup on the null hypothesis be altered ?

Which of the two designs would you prefer and why ?

SOLUTION. (*a*) *When all the twelve cups are presented to the lady in random order.*

According to the fundamental rule of permutation, the total number of ways in which the 12 cups of tea, 6 made in one way, and 6 in the other way can be presented to the lady.

are $\frac{12!}{6!6!}$ or 924

Since it is known to the lady that 6 are of each kind, she must have any one of the above arrangements of the 12 cups to be presented to her.

Hence the desired probability is $\frac{1}{924}$

(*b*) *When 12 cups are presented in 6 pairs to the lady.*

Since each pair consist of either kind of tea preparation, the probability that the lady would correctly judge each pair is $\frac{1}{2}$.

As there are 6 pairs and each event is independent of the other, the compound probability that the lady would judge all the six pairs correctly is $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{64}$

PROBLEM 11.39. *In a given race the odds in favour of three horses A, B, C, are 7:3, 2:3 and 2:5 respectively. Assuming that a dead heat is impossible, find the chance that one of them will win the race.*

SOLUTION. Probability that A wins, or $p_1=\frac{1}{4}$

Probability that B wins, or $p_2=\frac{2}{5}$

Probability that C wins, or $p_3=\frac{2}{7}$

Since these are mutually exclusive events, the chance that one of them will win the race.

$$=p_1+p_2+p_3$$

$$=\frac{1}{4}+\frac{2}{5}+\frac{2}{7}=\frac{131}{140}$$

PROBLEM 11.40. *Find the probability of throwing 6 at least once in four throws with single die.*

SOLUTION. The probability of throwing 6 in one throw of a die or $p=\frac{1}{6}$.

And the probability of not throwing a 6 or $q=\frac{5}{6}$.

The probability of throwing 6 at least once in 4 throws of a single die is equal to the total of the probabilities of throwing 6 in all the 4 throws or in 3 out of 4 throws or in 2 out of 4 throws or in one out of 4 throws. These probabilities can be found out by substituting values in the following equation.

$p={}^{n}C_{r}p^{r}q^{n-r}$ or we can expand the following equation.

$(p+q)^4$ or $(\frac{1}{6}+\frac{5}{6})^4$

$=p^4+4p^3q+6p^2q^2+4pq^3+q^4$

$$=\frac{1}{1296}+\frac{20}{1296}+\frac{150}{1296}+\frac{500}{1296}+\frac{625}{1296}=\frac{1296}{1296}=1$$

To find out the probability of throwing 6 at least once in 4 throws of a single die we must add the first 4 values because in the last value (q^4) there is no success at all. The required probability

therefore$=\frac{1}{1296}+\frac{20}{1296}+\frac{150}{1296}+\frac{500}{1296}=\frac{671}{1296}$

PROBLEM 11.41. *Find the chance of throwing 9 at least, in a single throw with two dice.*

SOLUTION. 9 can be made up in 4 ways, with two dice and thus the chance of throwing 9 is$=\frac{4}{36}$

10 can be made up in 3 ways with two dice and thus the chance of throwing 10 is$=\frac{3}{36}$

11 can be made up in 2 ways with two dice and thus the chance of throwing 11 is $\frac{2}{36}$

12 can be made up in 1 way with two dice and thus the chance of throwing 12 is $\frac{1}{36}$

Since these events are mutually exclusive, the chance of throwing a number not less than 9 is the sum of these separate chances ;

Therefore the required chance $=\frac{4+3+2+1}{36}=\frac{5}{18}$

PROBLEM 11.42. *From 30 tickets marked with the first 30 numerals, one is drawn at random ; find the chance that,*

(*a*) *it is a multiple of 5 or of 7.* (*b*) *it is a multiple of 3 or of 7.*

SOLUTION. (*a*) The chance that the number is a multiple of 5 is $\frac{6}{30}$ and the chance that it is a multiple of 7 is $\frac{4}{30}$ and these events are mutually exclusive, hence the required chance is $\frac{6}{30}+\frac{1}{30}=\frac{7}{30}$

(*b*) The chance that the number is a multiple of 3 is $\frac{10}{30}$ and the chance that is is multiple of seven is $\frac{4}{30}$ but 21 is common multiple of both three and seven. Hence the probability that it is either a multiple of 3 or of 7 is equal to $\frac{10}{30}+\frac{4}{30}-\frac{1}{30}=\frac{13}{30}$

PROBLEM 11.43. *A bag contains 5 white and 3 black balls, and are successively drawn out and not replaced, what is the chance that they are alternatively of different colour ?*

SOLUTION. (*a*) *Beginning with white.*

the probability of drawing a white $=\frac{5}{8}$

the probability of drawing a black then $=\frac{3}{7}$

the probability of drawing a white then $=\frac{4}{6}$

the probability of drawing a black then $=\frac{2}{5}$

therefore the probability of the compound event ;

$$=\tfrac{5}{8}\times\tfrac{3}{7}\times\tfrac{4}{6}\times\tfrac{2}{5}=\frac{1}{14}$$

(*q*) *Similarly beginning with black.*

the probability of drawing a black =[illegible]
the probability of drawing a white then =[illegible]
the probability of drawing a black then $=2$
the probability of drawing a white then $=\tfrac{4}{5}$

Therefore the probability of the compound event

$$=\tfrac{3}{8}\times\tfrac{5}{7}\times\tfrac{2}{6}\times\tfrac{4}{5}=\frac{1}{14}$$

The above two events are mutually exclusive. Therefore the required chance that 4 successive drawn balls are alternatively of different colour (without mentioning the colour with which to begin.)

$$=\frac{1}{14}+\frac{1}{14}=\frac{1}{7}$$

PROBLEM 3.44. *A and B throw alternately with a pair of dice. A wins if he throws 6 before B throws 7 and B if he throws 7 before A throws 6.*

If A begins, show that his chance of winning is $\frac{30}{61}$

SOLUTION. Chance of throwing 6 with 2 dice is $\frac{5}{36}$

and chance of throwing 7 with 2 dice is $\frac{6}{36}=\frac{1}{6}$

If A is to win, he should throw 6 on either the first, third or fifth..............throw.

Probability of A's throwing 6 on the 1st throw or $p_1=\frac{5}{36}$

Probability of A's throwing 6 on the 3rd throw or p_2

$$=\left[1-\left\{\left(1-\frac{5}{36}\right)\left(\frac{1}{6}\right)\left(\frac{5}{36}\right)\right\}\right]\times\frac{5}{36}=\frac{775}{7776}$$

Similarly, probability of A's throwing 6 on the 5th throw or p_3

$$=\frac{120125}{1679616}$$

and so on.

The sum of these probabilities which appear as the following geometrical progressions is A's chance of winning.

$$=\frac{5}{36}+\frac{775}{7776}+\frac{120125}{1679616}+\ldots\ldots\ldots\ldots$$

$$=\frac{\frac{5}{36}}{1-\left(\frac{775}{7776}\div\frac{5}{36}\right)}\qquad=\frac{\frac{5}{36}}{1-\frac{155}{216}}$$

$$=\frac{5}{36}\times\frac{216}{16}=\frac{30}{61}.$$

PROBLEM 11.45. *A, B and C in order toss a coin. The first one who throws a head wins. What are their respective chances ?*

SOLUTION. If A is to win, a head must be thrown on either the first, urth, or seventh, toss and of no other. The sum of these probabilities is A's ce of winning. Similarly, if B is to win a head must appear only on the ond, fifth, eighth........toss ; and if C is to win, the head must appear on the ird, sixth, ninth......toss.

These respective probabilities then appear as the following geometrical gressions :

$$A=\frac{1}{2}\div\frac{1}{16}+\frac{1}{128}+\dots\;\dots\dots=\frac{\frac{1}{2}}{1-(\frac{1}{16}\div\frac{1}{2})}=\frac{\frac{1}{2}}{1-\frac{1}{8}}=\frac{4}{7}$$

$$B=\frac{1}{4}\div\frac{1}{32}+\frac{1}{256}+\dots\dots\dots=\frac{\frac{1}{4}}{1-(\frac{1}{32}\div\frac{1}{4})}=\frac{\frac{1}{4}}{1-\frac{1}{8}}=\frac{2}{7}$$

$$C=\frac{1}{8}\div\frac{1}{64}+\frac{1}{512}+\dots\dots\dots=\frac{\frac{1}{8}}{1-(\frac{1}{64}\div\frac{1}{8})}=\frac{\frac{1}{8}}{1-\frac{1}{8}}=\frac{1}{7}$$

Thus the respective chances of winning of A, B and C are 4 : 2 : 1.

PROBLEM 11.46. *A, B and C, in order draw from a pack of cards, replacing eir card after each draw. If the first man to draw a heart wins, what are their spective chances ?*

SOLUTION. If A is to win, heart should be drawn on either the first, fourth r seventh.....draw and on no other. The sum of these probabilities is A's chance f winning.

Similarly if B is to win, a heart should appear only on the second, fifth, ghth... draw and on no other ; if C is to win, the heart must appear on the ird, sixth, ninth........toss.

These respective probabilities then appear as the following geometrical ogressions :

$$A:\frac{1}{4}+\frac{27}{256}+\frac{729}{16348}+\dots\dots$$

$$=\frac{1/4}{1-\left(\frac{27}{256}\div\frac{1}{4}\right)}=\frac{1/4}{1-\frac{27}{64}}=\frac{16}{37}$$

$$B:\frac{3}{16}+\frac{81}{1024}+\frac{2187}{65536}+\dots\dots$$

$$=\frac{3/16}{1=\left(\frac{81}{1024}\div\frac{3}{16}\right)}=\frac{\frac{3}{16}}{1-\frac{27}{64}}=\frac{12}{37}$$

$$C:\frac{9}{64}+\frac{243}{4096}+\frac{6561}{262144}+\dots\dots$$

$$=\frac{9/64}{1-\left(\frac{243}{4096}\div\frac{9}{64}\right)}=\frac{9/64}{1-\frac{2}{64}}=\frac{9}{37}$$

Thus, the respective chances of A, B and C to draw a heart first are : 12 : 9.

PROBLEM 17.47. (*a*) *Given n independent events with respective probab of occurrence,*

$$p_1 ; p_2 \ldots\ldots\ldots\ldots\ldots pn.$$

Write down the probability of at least one of these events happening.

(*b*) *What is the probability of getting 9 cards of the same suit in one hand a game of bridge ?*

SOLUTION. (*a*) Chance that all the events fail is,

$$(1-p_1)\ (1-p_2)\ldots\ldots\ldots\ldots(1-pn).$$

and except in this case some one of the events must happen ;

Hence, the required chance :

$$=1-(1-p_1)\ (1-p_2)\ldots\ldots\ldots\ldots(1-pn)$$

(*b*) The probability of getting 9 cards of the same suit in one hand at a of bridge

$$=\frac{{}^{13}C_9\times{}^{39}C_4}{{}^{52}C_{13}}\times 4$$

PROBLEM 11.48. *Often balls in a bag, three are red.*

(*a*) *What is the probability that there will be at least one red ball in a d of two balls ?*

(*b*) *What is the probability that both will be red ?*

SOLUTION. (*a*) *probability that at least one red ball will be drawn in draw of 2*

The number of ways in which 2 balls may be drawn is ${}^{10}C_2$

The number of ways in which 2 not red balls may be drawn is ${}^{7}C_2$

So, the probability that no red ball is drawn in a draw of 2.

$${}^{7}C_2/{}^{10}C_2=\frac{7\times 6}{10\times 9}=\frac{7}{15}$$

Therefore the probability that there will be at least one red ball in a d of two balls.

$$=1-\frac{7}{15}=\frac{8}{15}$$

(*b*) *probability that both balls will be red.*

The number of ways in which 2 red balls may be drawn out of 3, is ${}^{3}C_2$ and the number of ways in which 2 balls may be drawn is ${}^{10}C_2$, therefore required probability,

$$={}^{3}C_2/{}^{10}C_2=\frac{3\times 2}{10\times 9}=\frac{1}{15}$$

PROBLEM 11.49. *It is 8:5 against a person who is now 40 years old till he is 70 and 4 : 3 against a person now 50 living till he is 80. Find the bability that one at least of these persons will be alive 30 years hence.*

SOLUTION. Probability that the first person dies is $8/13$.

Probability that the second person dies is $4/7$.

Therefore, the probability of the compound event that both the persons not living 30 years hence,

$$=\frac{8\times 4}{13\times 7}=\frac{32}{91}$$

Hence the probability that one at least of these persons will be alive years hence.

$$=1-\frac{32}{91}=\frac{59}{91}$$

PROBLEM 11.50. *The odds against a certain event are 5 to 3 and the odds vour of another event independent of the former are 7 to 5 ; find the chance that at least of the events will happen.*

SOLUTION Chance that the first event does not happen is $\frac{5}{8}$

Chance that the second event does not happen is $1-\frac{7}{12}=\frac{5}{12}$

Therefore, the chance of the compound event that both the aforesaid events do not happen.

$$=\frac{5}{8}\times\frac{5}{12}=\frac{25}{96}$$

Hence, the chance that one at least of the events will happen.

$$=1-\frac{25}{96}=\frac{71}{96}$$

PROBLEM 11.51. *Five letters are placed at random in five addressed velopes. Find the chance that all the letters are not despatched in the right velopes.*

SOLUTION. The total number of ways of placing 5 letters in 5 envelopes !, all the letters can be sent correctly in one way only ;

nce its chance $=\frac{1}{5\,!}=\frac{1}{120}$

Therefore the chance that they are not all sent correctly

$$=1-\frac{1}{120}=\frac{119}{120}$$

PROBLEM 11.52. *There are four letters and four addressed envelopes. If letters are put at random in the envelopes, find the chance that no letter will cupy its correct place.*

SOLUTION. Four letters can be placed in four envelopes in 4 ! or 24 ways.

The first letter can be placed in a wrong envelope in 3 ways ✓

The second letter can be placed in a wrong envelope in 3 ways ✓
(Because the first letter can occupy the envelope meant for the second.)

The third letter can be placed in a wrong envelope in 2 ways ✓
(Because the first and second letters can occupy envelopes meant for 2nd and 3rd letters)

The fourth letter can now be placed in a wrong envelope in 1 way ✓

Total number of ways of putting all letters in wrong envelopes is 9

∴The required probability $=\frac{9}{24}$ ✓

In general the number of derangements of n things

$$=n\left\{1-\frac{1}{1\,!}\div\frac{1}{2\,!}+\frac{1}{3\,!}+\frac{1}{4\,!}\;\dots\dots+(n-1)\frac{1}{n\,!}\right\}$$

In the above case

$$=4!\left\{\frac{1}{2\,!}-\frac{1}{3\,!}+\frac{1}{4\,!}\right\}=24\times\left\{\frac{1}{2}-\frac{1}{6}+\frac{1}{24}\right\}$$

$$=24\times\frac{9}{24}=9$$

Hence the probability of a derangement $=\frac{9}{24}$

PROBLEM 11.53. *An illiterate person puts cardboard letters A, B, C and E in a line. What is the probability that all letters are at wrong places ?*

SOLUTION. Number of derangements of 5 things

$$= 5\,!\left\{\frac{1}{2\,!}-\frac{1}{3\,!}+\frac{1}{4\,!}-\frac{1}{5\,!}\right\}=44$$

The total number of ways of arranging 5 things $= 5\,! = 120$

$\therefore$ The required probability $= \frac{44}{120} = \frac{11}{30}$

✓ **PROBLEM 11.54.** *A problem in statistics is given to three studen R. G. T. whose chances of solving it are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$, What is the probability th the problem will be solved ?*

SOLUTION.

Chance that R fails to solve the problem is $1-\frac{1}{2}$, or $\frac{1}{2}$;
Chance that G ,, ,, ,, ,, is $1-\frac{1}{3}$, or $\frac{2}{3}$;
and the Chance that T ,, ,, ,, is $1-\frac{1}{4}$, or $\frac{3}{4}$;

Hence the probability that all the three students fail to solve the probl
$= \frac{1}{2}\times\frac{1}{3}\times\frac{1}{4} = \frac{1}{4}$.

Therefore the probability that the problem will be solved
$= 1-\frac{1}{4} = \frac{3}{4}$.

PROBLEM 11.55. *A has six shares in a lottery in which there are thr prizes and ten blanks. B has two shares in a lottery in which there are four priz and eight blanks. Which has the better chance to win a prize ?*

SOLUTION. A will draw a prize unless all of his six shares are blank Six tickets can be chosen from the thirteen in ${}^{13}C_6$ or 1716 ways. But, six blan can be drawn in ${}^{10}C_6$, or 210 ways. Hence the probability that A will draw bla is $\frac{210}{1716}$ ways and therefore the probability that he will draw at least one pr is $1-\frac{210}{1716} = \frac{1506}{1716} = \frac{251}{286}$

Similarly, the probability that B will win a prize is

$$1-\frac{{}^{8}C_2}{{}^{12}C_2} = 1-\frac{14}{33} = \frac{19}{33}.$$

Since the first value is larger than the second, it is seen that A has a bette chance than B to win a prize.

PROBLEM 11.56. *A has four shares in a lottery in which there are f prizes and five blanks. B has three shares in which there are three prizes and f blanks. Which has the better chance of winning exactly one prize ? Which winning two prizes ?*

SOLUTION. (*a*) *Chance of winning exactly one prize*

Chance that A has exactly 1 prize share

$$= {}^{4}C_1\times{}^{5}C_3/{}^{9}C_4 = \frac{4\times5\times4}{2}\times\frac{4\times3\times2\times1}{9\times8\times7\times6} = \frac{20}{63}$$

Chance that B has exactly one prize share

$$= {}^{3}C_1\times{}^{4}C_2/{}^{7}C_3 = \frac{3\times4\times3}{2}\times\frac{3\times2\times1}{7\times6\times5} = \frac{18}{35}$$

Since the second value is larger than the first it is seen that B has bette chances than A of winning exactly one prize.

(*b*) *Chance of winning two prizes,*

Chance that A has exactly 2 prize share.

$$= {}^{4}C_2\times{}^{5}C_2/{}^{9}C_4 = \frac{4\times3}{2\times1}\times\frac{5\times4}{2\times1}\times\frac{4\times3\times2\times1}{9\times8\times7\times6} = \frac{10}{21}$$

Chance that B has exactly two prize shares

$= {}^3C_2 \times {}^4C_1 / {}^7C_3 = \frac{3\times4\times3\times2\times1}{7\times6\times5} = \frac{12}{35}$

Since the first value is larger than the second, it is seen that A has better chances than B of winning two prizes.

PROBLEM 11.57. *A can hit a target 3 times in 5 shots, B 2 times in 5 shots, C 3 times in 4 shots. They fire a volley. What is the probability that 2 shots hit ?*

SOLUTION. In order that two shots may hit the target, the following mutually exclusive cases must be considered.

Chance that A and B hit and C fails to hit target, or $p_1 = \frac{3}{5} \times \frac{2}{5} \times (1-\frac{3}{4}) = \frac{6}{100}$

Chance that A and C hit and B fails to hit the target or $p_2 = \frac{3}{5} \times \frac{3}{4} \times (1-\frac{2}{5}) = \frac{27}{100}$

Chance that B and C hit and A fails to hit the target or $p_3 = \frac{2}{5} \times \frac{3}{4} \times (1-\frac{3}{5}) = \frac{12}{100}$

Since they are mutually exclusive events, the probability that any 2 shots hit $= p_1 + p_2 + p_3$

$= \frac{6}{100} + \frac{27}{100} + \frac{12}{100} = \frac{45}{100} = \frac{9}{20}$

PROBLEM 11.58. *The odds that a book will be favourably reviewed by the three independent critics are 3 to 2, 4 to 3, and 2 to 3 respectively. What is the probability that of the three reviews a majority will be favourable ?*

SOLUTION. In order a majority of the 3 reviews is favourable, two must be favourable.

Probability that the first two are favourable and the third unfavourable, or $p_1 = \frac{3}{5} \times \frac{4}{7} \times (1-\frac{2}{5}) = \frac{36}{175}$

Probability that the first and the third are favourable and the second unfavourable, or $p_2 = \frac{3}{5} \times \frac{2}{5}(1-\frac{4}{7}) = \frac{18}{175}$

Probability that the last two are favourable and the first unfavourable, or $p_3 = \frac{4}{7} \times \frac{2}{5} \times (1-\frac{3}{5}) = \frac{16}{175}$

If any one of the above three combinations happens, the book can be said to have been reviewed favourably by a majority.

Since they are mutually exclusive events, the chance that the event will happen

$$= p_1 + p_2 + p_3 = \frac{36}{175} + \frac{18}{175} + \frac{16}{175}$$
$$= \frac{70}{175} = \frac{2}{5}$$

PROBLEM 11.59. *A is one of five horses entered for a race and is to be ridden by one of two jockeys B and C ; it is 3 to 1 that B rides A, in which case all the horses are equally likely to win, if C rides A, his chance is doubled. What are the odds in favour of his winning ?*

SOLUTION. Chance that [A] is ridden by B and wins the race or p_1 $= \frac{3}{4} \times \frac{1}{5} = \frac{3}{20}$

Chance that [A] is ridden by C and wins the race or p_2

$= \frac{1}{4} \times \frac{2}{5} = \frac{2}{20}$

Since the given event that A wins the race can happen in two different ways as given above, the chance that it will happen is

$= p_1 + p_2$

$= \frac{3}{20} + \frac{2}{20} = \frac{5}{20} = \frac{1}{4}$

Therefore the odds in favour of A's winning the race are 1 to 3.

PROBLEM 11.60 *Fourteen quarters and one five-dollar gold piece are in one purse and fifteen quarters are in another. Ten coins are taken from the second an put into the first. Which purse is probably the more valuable ?*

SOLUTION. The purse containing the five-dollar gold piece would be more valuable. We have therefore to compare the probability that it is in the first purse with the probability that it is in the second. The probability that the gold piece was taken from the first purse and put into the second is $\frac{10}{15}$ and the probability that, being in second, it was again returned to the first is $\frac{10}{25}$

Hence the Probability the gold piece was transferred from the first purse to the second purse and remained there.

$$=\left(\frac{10}{15}\right)\left(1-\frac{10}{25}\right)=\frac{2}{5}$$

The first purse has a larger value than the second, because the probability that the gold coin remained in the second purse is only.

PROBLEM 11.61. *Suppose that one urn A contains 4 black balls, and 5 white balls, while a second urn B contains 3 black and 7 white balls. Find out the probability that a blind-folded person in one draw shall obtain a white ball.*

SOLUTION. Probability that the blind person chooses the urn A and draws a white ball$=\frac{1}{2}\times\frac{5}{9}=\frac{5}{18}$.

Probability that the blind person chooses the urn B and draws a white ball $=\frac{1}{2}\times\frac{7}{10}=\frac{7}{20}$

Since the event, that is, the draw of white ball by a blind person can happen in the above two different ways ways which are mutually exclusive, the chance that the event will happen is

$=\frac{5}{18}+\frac{7}{20}=\frac{113}{180}$

PROBLEM 11.62. *What is the chance of drawing a one-rupee coin from a purse, one compartment of which contains 2 one-rupee and 4 Twenty five paisa coins ; the second 10 one rupee coins ; the third 10 Twenty five paisa coins and 5 one paisa coins and the fourth 7 one-rupee and 5 twenty Five paisa coins.*

SOLUTION Chance that the first compartment of the purse is chosen and a one rupee coin is drawn or $p_1=\frac{1}{4}+\frac{2}{6}=\frac{1}{12}$

Chance that the second compartment is chosen and a one-rupee coin is drawn or $p_2=\frac{1}{4}\times\frac{10}{10}=\frac{1}{4}$

Chance that third compartment is chosen and a one-rupee coin is drawn or $p_3=\frac{1}{4}+\frac{0}{15}=0$.

Chance that the fourth compartment is chosen and a one-rupee coin is drawn or $p_4=\frac{1}{4}\times\frac{7}{12}=\frac{7}{48}$

Since the event, that is, the draw of a one-rupee coin from a purse can happen in the above four different ways, the chance that the one-rupee coin will be drawn from the purse is

$=p_1+p_2+p_3+p_4$
$=\frac{1}{12}+\frac{1}{4}+0+\frac{7}{48}=\frac{23}{48}$

PROBLEM 11.63. *A man draws from an urn containing two balls, one white and one black. If he draws a white ball he wins. If he fails to draw a white ball, the draw is replaced, another black ball is added and he draws again. If he fails to draw a white ball in the next draw, the process is repeated. What are his respective chances of winning in 2, 3, 4, 5, 6, 7, and 10 trials ?*

SOLUTION. Chance of winning

First Trial $p=\frac{1}{2}$
Second Trial $p=(1-\frac{1}{2})\times\frac{1}{3}=\frac{1}{6}$
Third Trial $p=\{1-(\frac{1}{2}+\frac{1}{6})\}\times\frac{1}{4}=\frac{1}{12}$
Fourth Trial $p=\{1-(\frac{1}{2}+\frac{1}{6}+\frac{1}{12})\}\frac{1}{5}=\frac{1}{20}$
Fifth Trial $p=\{1-(\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\frac{1}{20})\}\frac{1}{6}=\frac{1}{30}$
Sixth Trial $p=\{1-(\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\frac{1}{20}+\frac{1}{30})\}\frac{1}{7}=\frac{1}{42}$
Seventh Trial $p=\{1-(\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\frac{1}{20}+\frac{1}{30}+\frac{1}{42})\}\frac{1}{8}=\frac{1}{56}$

Eighth Trial $p=\{1-(\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\frac{1}{20}+\frac{1}{30}+\frac{1}{42}+\frac{1}{56})\}\frac{1}{9}=\frac{1}{72}$

Ninth Trial $p=\{1-(\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\frac{1}{20}+\frac{1}{30}+\frac{1}{42}+\frac{1}{56}+\frac{1}{72})\}\frac{1}{10}=\frac{1}{90}$

Tenth Trial $p=\{1-(\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\frac{1}{20}+\frac{1}{30}+\frac{1}{42}+\frac{1}{56}+\frac{1}{72}+\frac{1}{90})\frac{1}{11}=\frac{1}{110}$

PROBLEM 11.64. *In a certain game A's skill is to B's as 3 is to 2 ; find the chance of A winning 3 games out of 5.*

SOLUTION. Since p = A's chance of winning in one game $=\frac{3}{5}$ and $q=1-\frac{3}{5}=\frac{2}{5}$

Therefore the probability of A's winning 3 games out of 5 = $^5C_3p^3q^2$

$$=\frac{5\times4}{2\times1}\times\frac{3}{5}\times\frac{3}{5}\times\frac{3}{5}\times\frac{2}{5}\times\frac{2}{5}=\frac{216}{625}$$

PROBLEM 11.65. *In each of a set of games it is 2 to 1 in favour of the winner of the previous game, what is the chance that the player who wins the first game shall win at least three of the next four ?*

SOLUTION The player who has won the 1st game has the following chance of winning or losing the other games :—

(*a*) Of losing the 1st game and winning the remaining three $\frac{1}{3}\times\frac{1}{3}\times\frac{2}{3}\times\frac{2}{3}$

(*b*) ,, ,, 2nd ,, ,, ,, ,, $\frac{2}{3}\times\frac{1}{3}\times\frac{1}{3}\times\frac{2}{3}$

(*c*) ,, ,, 3rd ,, ,, ,, ,, $\frac{2}{3}\times\frac{2}{3}\times\frac{1}{3}\times\frac{1}{3}$

(*d*) ,, ,, 4th ,, ,, ,, ,, $\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}\times\frac{1}{3}$

(*e*) Of all the four games ,, ,, ,, ,, $\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}$

Since these events are mutually exclusive therefore the probability of the event happening is equal to the sum of the probabilities of the above 5 events.

It is equal to $\frac{4}{81}+\frac{4}{81}+\frac{4}{81}+\frac{8}{81}+\frac{16}{81}=\frac{36}{81}=\frac{4}{9}$

PROBLEM 11.66. *If the chance that a vessel arrives safely at a port is* $\frac{9}{10}$ *find the chance that out of 5 vessels expected 4 at least will arrive safely.*

SOLUTION. In the case of a single vessel the chance of safe arrival is $\frac{9}{10}$;

and the chance of unsafe arrival is $\frac{1}{10}$

Now the required event follows if 4 ships or 5 ships arrive safely ; therefore the required chance is the sum of first two terms of the expansion of

$$\left(\frac{9}{10}+\frac{1}{10}\right)^5$$

Thus the chance $=\frac{1}{10^5}(9^5+5\times9^4\times1)=\frac{59049+(5\times6561\times1)}{100000}=\frac{91854}{100000}$

$$=\frac{45927}{50000}$$

PROBLEM 11.67. *Three coins are tossed simultaneously. What is the probability that they will fall two heads and one tail ?*

SOLUTION. p, *i. e.*, the chance that a coin falls head is $\frac{1}{2}$ and so $q=1-\frac{1}{2}=\frac{1}{2}$.

Now, the probability out of 3 coins exactly 2 fall heads and one tail

$=^3C_2p^2q=3\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{3}{8}$

PROBLEM 11.68 *Five coins whose faces are marked 2, 3 are thrown. What is the chance of obtaining a total of 12 ?*

SOLUTON. To make a total of 12, two coins must fall with the mark 3 up and three coins with the mark 2 up. No other arrangement would give a total of 12.

Now p. *i. e.*, the probability that a coin in one throw falls with the mark 3 up $= \frac{1}{2}$

and $q = 1 - \frac{1}{2} = \frac{1}{2}$

So, the chance of obtaining a total of 12 with 5 coins

$= {}^5C_2 p^2 q^3$

$= \frac{5\times4}{2\times1} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{5}{16}$

PROBLEM 11.69 *A coin whose faces are marked 3 and 5 is tossed 4 times What are the odds against the sum of the numbers thrown being less than 15 ?*

SOLUTION. To make a sum of less than 15, in 4 tosses, the mark 3 must appear at least 3 times.

Now the chance of the coin falling with the mark 3 up in one toss is $\frac{1}{2}$. and the chance of not coming with the mark 3 is also $\frac{1}{2}$!. Now the required event follows if the mark 3 appears three times or four times therefore the required chance is the sum of the first two terms of the expansion of $(\frac{1}{2}+\frac{1}{2})^4$

Thus the chance of numbers thrown being less than 15

$= (\frac{1}{2})^4 + (4 \times (\frac{1}{2})^3 \times \frac{1}{2}) = \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$

Therefore the odds against the sum of the numbers thrown being less than 15 are 1 to 5.

PROBLEM 11.70. *From a bag containing 24 quarters and one five-dollar gold piece, a person is allowed to draw 5 coins indiscriminately; find the value of his expectations.*

SOLUTION. There are 25 coins valuing 44 quarters out of which he can draw 5.

Therefore the value of his expectation $= \left(\frac{5}{25} \times 44\right)$ quarters

$= 8\frac{4}{5}$ quarters

PROBLEM 11.71 *There are 9 notes & coins in a bag, 6 of which are 2 rupee notes and the rest are unknown coins of equal values ; find what they must be if the probable value of a draw is Rs. 1.50*

SOLUTION. Since the probable value of draw is Rs. 1.50,bag having 9 notes & coins values $(1.50\times9)=13.50$ rupees. Out of this there are 6,2 rupees notes and 3 unknown.

The value of 3 unknown coins = Rs. 1.50

or the value of 1 unknown coin = 50 paise coin

Thus there are 3,50 paisa coins in the bag.

PROBLEM 11.72 *A pays B 1 rupee to guess the number of heads in a single toss of 4 coins; what expectation should B place on each of the possibilities; no head, one head, etc. ?*

SOLUTION. (1) Probability of no head is $\frac{1}{2^4} = \frac{1}{16}$

$\therefore$ B's expectation on no head $= \frac{1}{16} \times 1$ rupee. = .0625 rupee

(2) Probability of one head $= {}^4C_1 pq^3$, where $p = q = \frac{1}{2}$.

$= 4 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{4}{16}$

$\therefore$ B's expectation on one head $= \frac{4}{16} \times 1$ rupee$=.25$ rupee

(3) Probability of two heads $= {}^4C_2 p^2 q^2$, where $P = q = \frac{1}{2}$

$$\frac{4\times 3}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} = \frac{3}{8}$$

B's expectation on two heads $= \frac{3}{8} \times 1$Re. $= .375$ rupee.

PROBLEM 11.73. *A and B throw with one die for a prize of Rs. 11, which is to be won by the player who first throws 6. If A has the first throw, that are their respective expectations.*

SOLUTION. If A is to win, six must be thrown on first, thiro, sixth.... ......throws and no other. The sum of these probabilities is A's chance of winning. Similarly if B is to win, six must be thrown on second, fourth, sixth.... .....throws.

These respective probabilities then appear as following geometrical progressions.

$$\text{A} :\!-\; \frac{1}{6} + \frac{25}{216} + \cdots\cdots$$

$$= \frac{1/6}{1 - \left(\frac{25}{216} \div \frac{1}{6}\right)} = \frac{1/6}{1 - \frac{25}{36}} = \frac{6}{11}$$

$$\text{B} :\!-\; \frac{5}{36} + \frac{125}{1296} \cdots\cdots$$

$$= \frac{5/36}{1 = \left(\frac{126}{1296} \div \frac{5}{36}\right)} = \frac{5/36}{1 - \frac{25}{36}} = \frac{5}{11}$$

Therefore for a prize of Rs. 11.

A's expectation $= \frac{6}{11} \times$ Rs. $11 =$ Rs. 6.

and B's expectation $= \frac{5}{11} \times$ Rs. $11 =$ Rs. 5.

PROBLEM 11.74. *Suppose a black ball has been drawn from one of three bags, the first containing three black balls and seven white, the second five black balls and three white, the third eight black balls and four white; what is the probability that it was drawn from the first bag ?*

SOLUTION. If an event is known to have proceeded from one of n mutually exclusive causes whose probabilities are $P_1, P_2 \ldots\ldots P_n$ and furthermore it $p_1, p_2 \ldots\ldots, p_n$ are the respective probabilities that one of the n causes exists, then the probability that the event proceeded from the mth cause is

$$P = \frac{Pm.Pm}{P_1 p_1 + P_2 p_2 + \cdots\cdots + P_n p_n}$$

In the given problem

$P_1 = P_2 = P_3 = \frac{1}{3}$, since it is just as probable that the ball was drawn from one bag as another.

also, p_1 *i.e.*, the probability of drawing a black ball from the first bag$= \frac{3}{10}$

and p_2, *i. e.*, the probability of drawing a black ball from the second bag$=\frac{5}{8}$.

and p_3, *i.e.* ,, ,, ,, ,, ,, ,, ,, third bag $= \frac{8}{12}$

Therefore, the probability that a black ball drawn belongs to the first bag.

$$=\frac{1/3\,.\,3/10}{1/3.3/10+1/3.5/8+1/3.8/12}=\frac{36}{191}$$

PROBLEM 11.75 *A bag contains 5 balls which are just as likely to be white as coloured. Two white balls are drawn from the bag. What is the probability that all are white ?*

SOLUTION. The probability that a ball is white $=\frac{1}{2}$

The probability that all the balls are white$={}^5C_5p^5=\frac{1}{3}$

,, ,, ,, 4 ,, ,, ,, $={}^5C_4p^4q=\frac{5}{32}$

,, ,, ,, 3 ,, ,, ,, $={}^5C_3p^3q=\frac{10}{32}$

,, ,, ,, 2 ,, ,, ,, $={}^5C_2p^2q=\frac{10}{32}$

Two white balls have been drawn.

The probability of drawing 2 white balls when all the balls are white

$$=\frac{{}^5C_2}{{}^5C_2}=1$$

,, ,, , 2 ,, ,, 4 ,, $=\frac{{}^4C_2}{{}^5C_2}=\frac{6}{10}$

,, ,, ,, 2 ,, ,, 3 ,, $=\frac{{}^3C_2}{.{}^5C_2}=\frac{3}{10}$

,, ,, ,, 2 ,, ,, 2 ,, $=\frac{{}^2C_2}{{}^5C_2}=\frac{1}{10}$

The probability of the compound event that all the balls were white and two white balls were drawn$=\frac{1}{32}\times 1=\frac{1}{32}$

The probability of the compound event that there were 4 white balls and two white balls were drawn $\frac{5}{32}\times\frac{6}{10}=\frac{30}{320}$

3 ,, ,, ,, ,, ,, $\frac{10}{32}\times\frac{3}{10}=\frac{30}{320}$

2 ,, ,, ,, ,, ,, $\frac{10}{32}\times\frac{1}{10}=\frac{10}{320}$

Since the events are mutually exclusive the total of the probability

$$p=\frac{1}{32}+\frac{30}{320}=\frac{30}{320}+\frac{10}{320}=\frac{80}{320}$$

If the total of probability $= \frac{80}{320}$ the probability of all balls being white $= \frac{1}{32}$

" " " " $=1$ " " " " $= \frac{1}{32} \times \frac{320}{80}$

$= \frac{1}{8}$

PROBLEM 11.76 *Suppose that A is known to tell the truth in five cases out of six and he states that a white ball was drawn from a bag containing 9 black and one white ball ; what is the probability that the white ball was really drawn ?*

SOLUTION. The probability that a white ball is drawn in any case is 1/10. Also, the probability that the white ball was drawn and that A told the truth is $1/10 \times 5/6$.

Furthermore, the probability that a black ball was drawn and A told a lie about it is $9/10 \times 1/6$.

Hence the probability that a white ball was drawn is

$$p = \frac{1/10.5/6}{1/10.5/6 + 9/10.1/6} = 5/14$$

PROBLEM 11.77 *(a) The probability that a certain event happened was 1/10, and A, who is accurate in 49 cases out of 50, said that it happened. B agrees with A in stating that the event happened. B is accurate in 9 cases out of 10. What is the probability that it actually did occur ?*

(b) Suppose if C, who is accurate in 7 cases out of 10, denies that the event mentioned above happened, what is the probability that it happened ?

SOLUTION. (*a*) Probability that the event happened and that A and B were right.

$$= \frac{1}{10} \times \frac{49}{50} \times \frac{9}{10} = \frac{441}{5000}$$ ✓

Probability that the event did not happen and that A and B were wrong

$$= \frac{9}{10} \times \frac{1}{50} \times \frac{1}{10} = \frac{9}{5000}$$

∴ Probability that the event actually did occur

$$\frac{441/5000}{441/5000 + 9/5000} = \frac{4415000}{4505000} = \frac{441}{450} = \frac{49}{50}$$ ✓

(*b*) Probability that the event happened and A and B were right and C wrong

$$= \frac{1}{10} \times \frac{49}{50} \times \frac{9}{10} \times \frac{3}{10} = \frac{1323}{50,000}$$

Probability that the event did not happen and A and B were wrong and C right $= \frac{9}{10} \times \frac{1}{50} \times \frac{1}{10} \times \frac{7}{10} = \frac{63}{50,000}$

∴ Probability that the event actually did occur

$$= \frac{1323/50,000}{1323/50,000 + 63/50,000} = \frac{1323/50,000}{1386/50,000} = \frac{1323}{1386} = \frac{21}{22}$$

PROBLEM 11.78. *A, B and C run a race. The odds are 8 to 3 against A, 5 to 2 against B. Assuming a dead heat to be impossible, find the odds against C.*

SOLUTION. As the events in question are mutually exclusive the total of the possibilities of A's, B's and C's success must be equal to unity.

or $\frac{3}{11}+\frac{2}{7}$ + C's chance of success = 1.

$\therefore$ C's chance of success $= 1 - \left(\frac{3}{11}+\frac{2}{7}\right) = \frac{34}{77}$

Odds against C are 43 to 34.

PROBLEM 11.79. *I have 5 rupees and 3 five paise coins placed at random in a line. What is the chance that both the extreme coins are 5 paise coins ?*

SOLUTION. The chance that the one extreme coin is a 5 paise $= \frac{3}{8}$, and after the event happening the chance that the other extreme position is also held to be a 5 paise $= \frac{2}{7}$,

$\therefore$ The chance of the compound event

$$= \frac{3}{8}\times\frac{2}{7} = \frac{3}{28}$$

In such cases the possibilities of the compound event are given by :—

$$p = \frac{n(n-1)}{(m+n)(m+n-1)}$$

Where '*n*' stands for the unit whose chance of success has to be calculated and '*m*' the other event, substituting the figures :—

$$p = \frac{3(3-1)}{(m+n)(m+n-1)}$$

Where '*n*' stands for the unit whose chance of success has to be calculated and '*m*' the other event, substituting the figures :—

$$p = \frac{3(3-1)}{(5+3)(5+3-1)} = \frac{6}{56} = \frac{3}{28}$$

PROBLEM 11.80. *Three cards are drawn at random from a pack, find the chance that they are an ace, a king and a queen.*

SOLUTION. Three cards can be drawn from 52 in ${}^{52}C_2$ ways, and one ace, one king and one queen can be selected out of 4 aces, 4 kings, and 4 queens in ${}^4C_1\times{}^4C_1\times{}^4C_1$ ways.

Therefore required chance of the event

$$= \frac{{}^4C_1\times{}^4C_1\times{}^4C_1}{{}^{52}C_3} = \frac{4\times4\times4\times3\times2}{52\times51\times50} = \frac{16}{5525}$$

PROBLEM. 11.81 *The odds in favour of a certain event are 2 to 5 and the odds against another event independent of the former are 5 to 6. Find the chance that one at least of the events will happen.*

SOLUTION. The chance that the 1st event happens and the 2nd one does not happen

$$= \frac{2}{7}\times\frac{5}{11} = \frac{10}{77}$$

The chance that the 1st event does not happen and the 2nd happens.

$$= \frac{5}{7}\times\frac{6}{11} = \frac{30}{77}$$

The chance that both the events happen

$$= \frac{2}{7}\times\frac{6}{11} = \frac{12}{77}$$

The chance that one at least of the events will happen.

$$= \frac{10}{77}+\frac{30}{77}+\frac{12}{77} = \frac{52}{77}$$

the Prob that both
the events do not

$= \left(1-\frac{2}{7}\right)\left(1-\frac{6}{11}\right)$

$= \frac{5}{7}\cdot\frac{5}{11} =$

Prob that at
events

PROBLEM 11.82. *The odds against A solving a problem are 8 to 6, and the odds in favour of B solving the same problem are 14 to 10. What is the probability that if both of them try the problem it will be solved ?*

SOLUTION. The probability that A will fail to solve the problem $= \frac{8}{14}$

The probability that B will fail to solve the problem $= \frac{10}{24}$

$\therefore$ The compound probability of none of them solving the problem

$$= \frac{8}{14} \times \frac{10}{24} = \frac{5}{21}$$

$\therefore$ The probability that the problem is solved either by A or by B or by both of them $= 1 - \frac{5}{21} = \frac{16}{21}$

PROBLEM 11.83. *Four persons draw each a card from an ordinary pack, find the chance that no cards are of equal value.*

SOLUTION. The probability that the second would draw a card different in value from the one drawn by the first person $= \frac{48}{51}$, as there are 51 cards left and they contain 3 cards of the value as drawn by the first person.

The probability that the third person would draw a card different in value from those drawn by the first and second persons $= \frac{44}{50}$, as there are 50 cards left and they contain 6 cards of the same value as drawn either by A or B.

Similarly, the probability that the fourth person would draw a card different in value from those drawn by the previous three persons $= \frac{40}{49}$

The probability of the compound event

$$= \frac{48}{51} \times \frac{44}{50} \times \frac{40}{49} = \frac{2816}{4165}$$

PROBLEM 11.84. *A purse contains 1 rupee and 9 copper coins of the size, weight and design of a rupee. Another purse contains 10 such copper coins only. 9 coins are taken from the first purse and put in the second one and then 9 coins are taken from the second purse and put in the first one. Find the chance that the rupee is still in the first purse.*

SOLUTION. The probability that the rupee did not go out of the first purse at all when 9 coins are taken out $= \frac{{}^9C_8}{{}^{10}C_9} = \frac{1}{10}$

The probability that the rupee went out of the first purse and then came back to it $= \left(\frac{{}^9C_8 \times {}^1C_1}{{}^{10}C_9} \times \frac{{}^{18}C_8 \times {}^1C_1}{{}^{19}C_9}\right)$

$$= \frac{9}{10} \times \frac{9}{19} = \frac{81}{190}$$

If any of the two events happens the rupee would remain in the first purse.

So its chances are $\frac{1}{10} + \frac{81}{190} = \frac{100}{190} = \frac{10}{19}$

PROBLEM 11.85. *If six coins are tossed what is the chance the one and only one will turn up head.*

SOLUTION. The probability of coin falling with head or tail upwards is equal or $p = q = \frac{1}{2}$

The probability of only one head upward would be given by that term of binomial in which there is p^1q^5.

$$(p+q)^6 = p^6 + 6p^5q^1 + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6p^1q^5 + q^6$$

So the sixth term of the binomial $(\frac{1}{2}+\frac{1}{2})^6$ would give us the desi probability.

$$(\tfrac{1}{2}+\tfrac{1}{2})^6 \quad \frac{1}{64}+\frac{6}{64}+\frac{15}{64}+\frac{20}{64}+\frac{15}{64}+\frac{6}{64}+\frac{1}{64}$$

The probability of one head upward $= \frac{6}{64}$ or $\frac{3}{32}$

Without raising the binomial we can find out the probability by ${}^{n}C_{r}p_{r}q^{n-r}$ or ${}^{6}C_1p^1q^5 = 6\times\frac{1}{2}\times(\frac{1}{2})^5 = 6\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2} = \frac{6}{64} = \frac{3}{32}$

PROBLEM 11.86. *A, B and C, in order to cut from a pack of card replacing them after each cut. The first to cut a diamond is to win, what are the respective chances ?*

SOLUTION. A's chance of success $= \frac{1}{4}$

B's chance of success $= \frac{3}{4}\times\frac{1}{4} = \frac{3}{16}$

as B will get a chance only if A has failed in the first attempt.

C's chance of success $= \frac{3}{4}\times\frac{3}{4}\times\frac{1}{4} = \frac{9}{64}$

as C will get a chance if both A and B fail

So their chances are $\frac{1}{4} : \frac{3}{16} : \frac{9}{64}$

or 16 : 12 : 9.

The probabilities of their success thus would be $\frac{16}{37}, \frac{12}{37}, \frac{9}{37}$.

PROBLEM 11.87. *P and Q draw in order from a purse in which are thre sovereigns and four shillings. If coins are not replaced after the draws, what ar their respective chances of drawing a sovereign ?*

SOLUTION. P's chances of success in the first draw $= \frac{3}{7}$.

Q's chance of success in the first draw.

$$= \frac{4}{7}\times\frac{3}{6} = \frac{2}{7}$$

P's chance of success in the second draw

$$= \frac{4}{7}\times\frac{3}{6}\times\frac{3}{5} = \frac{6}{35}$$

Q's chance of success in the second draw

$$= \frac{4}{7}\times\frac{3}{6}\times\frac{2}{5}\times\frac{3}{4} = \frac{3}{35}$$

At this stage there are only 3 sovereigns in the purse, so if they have faile so far P is sure to get it. The probability of their falling tiil now and P's gettin it would be,

$$\frac{4}{7}\times\frac{3}{6}\times\frac{2}{5}\times\frac{1}{4}\times\frac{3}{5} = \frac{1}{35}$$

Thus the total of P's chances

$$= \frac{3}{7}+\frac{6}{35}+\frac{1}{35} = \frac{22}{35}$$

and total of Q's chances

$$= \frac{2}{7}+\frac{3}{35} = \frac{13}{35}$$

PROBLEM 11.88. *Five persons throw for a stake which is to be won by th first person who throws a head with a rupee. If they throw in succession what ar the chances of each of them.*

SOLUTION. 1st person's chance of success $= \frac{1}{2}$
2nd ,, ,, ,, ,, $= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
3rd ,, ,, ,, ,, $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$
4th ,, ,, ,, ,, $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
5th ,, ,, ,, ,, $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{32}$

Thus their respective chances of success are in the ratio of 16 : 8 : 4 : 2 : 1 and so their chances are $\frac{16}{31}, \frac{8}{31}, \frac{4}{31}, \frac{2}{31}, \frac{1}{31}$.

PROBLEM 11.89. *A and B who are players of equal skill leave a badminton game when A had secured 12 points and B 13. If the game was to finish at 15, and the winner was to get Rs. 32, what share each ought to take ?*

SOLUTION. A would have won if he

(*i*) Won the next three points, the probability of which

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

(*ii*) Lost the 1st point and won the next three the probability of which

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

(*iii*) Lost the second point and won the 1st, 3rd and 4th points the probability of which $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

(*iv*) Lost the third point and won the 1st, 2nd and 4th the probability of which $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

So his chance of success $= \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{5}{16}$

B would have won if he

(*i*) Won the next two points the probability of which

$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

(*ii*) Lost the 1st point and won the 2nd. and third the probability of which

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

(*iii*) Lost the 2nd point and won the 1st and third the probability of which

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{8}$$

(*iv*) Lost the first and second and won the 3rd and 4th points the probability of which $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

(*v*) Lost the 2nd and 3rd and won the 1st and 4th points the probability of which $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

(*vi*) Lost the 1st and 3rd and won the 2nd and 4th points the probability of which

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

So his chance of success $= \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{11}{16}$

Since their chances of success are $= \frac{5}{16}$ and $\frac{11}{16}$

They should share $\left(\frac{5}{16} \times 32\right)$ and $\left(\frac{11}{16} \times 32\right)$

or Rs. 10 and Rs. 22 respectively.

PROBLEM 11.90. *P has in his pocket one rupee and four 5 paise coins. He takes out two coins at random and promises to give one coin to R and the other to Q. What is the worth of Q's expectation.*

SOLUTION. The probability of P's drawing 2 five paise coins

$$= \frac{{}^4C_2}{{}^5C_2} = \frac{6}{10} = \frac{3}{5}$$

The probability of P's drawing 1 five paise coin and 1 rupee

$$= \frac{{}^4C_1 \times {}^1C_1}{{}^5C_2} = \frac{4}{10} = \frac{2}{5}$$

The chance that P drew I five paise coin and the rupee and that *Q* got the rupee $= \frac{2}{5} \times \frac{1}{2} = \frac{1}{5}$ and the value of his expectation.

$$= \frac{1}{5} \times 20 \text{ five paise coins} = 4 \text{ five paise coins}$$

The chance that P drew I five paise coin and the rupee and Q got the five paise coin.

$$= \frac{2}{5} \times \frac{1}{2} = \frac{1}{5} \text{ and the value of his expectation}$$

$$= \frac{1}{5} \times 1 \text{ five paise coin} = \frac{1}{5} \text{ five paise coin}$$

Thus the total worth of the expectation

$$= \frac{3}{4} + 4 + \frac{1}{5} = 4\frac{4}{5} \text{ five paise coins}$$

PROBLEM 11.91. *There are three bails in a bag, but it is not known of what colour they are ; one drawn from the bag is of red colour. What is the probability that all balls are red ?*

SOLUTION. The probability that all the balls, or two balls or only one ball is red is equal, or it is $\frac{1}{3}$.

The probability that all the three balls are red and that single draw would give a red ball $= \frac{1}{3} \times \frac{3}{3} = \frac{3}{9}$

The probability that two balls are red and that a single draw would give a red ball $= \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$

The probability that only one ball is red and that a single draw gives a red ball $= \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

∴ The probability that all the balls are red

$$= \frac{3/9}{3/9 + 2/9 + 1/9} = \frac{3}{9} \times \frac{9}{6} = \frac{1}{2}$$

PROBLEM 11.92. *There are 6 balls in a bag. Their colours are not known. A draw of 3 balls is made and all of them are white. What is the probability that no white ball is left in the bag ?*

SOLUTION. The probability that there are 6. 5, 4 or 3 white balls in the bag is equal or it is $\frac{1}{4}$.

The probability that there are six white balls and a draw of 3 would give 3 white balls.

$$= \frac{1}{4} \times \frac{{}^6C_3}{{}^6C_3} = \frac{20}{80}$$

The probability that there are 5 white balls and a draw of 3 would give 3 white balls.

$$=\tfrac{1}{4}\times\frac{{}^5C_3}{{}^6C_3}=\frac{10}{80}$$

The probability that there are 4 white balls and a draw of three would give 3 white balls

$$=\tfrac{1}{4}\times\frac{{}^4C_3}{{}^6C_3}=\frac{4}{80}$$

The probability that there are three white balls and a draw of three would give three white balls

$$=\tfrac{1}{4}\times\frac{{}^3C_3}{{}^6C_3}=\frac{1}{80}$$

The probability that only three balls are white.

$$=\frac{\frac{1}{80}}{\frac{20}{80}+\frac{10}{80}+\frac{4}{80}+\frac{1}{80}}$$

$$=\frac{1}{80}\times\frac{80}{35}=\frac{1}{35}$$

PROBLEM 11.93. *A bag contains 5 balls. Their colours are not known. A ball is taken out and found to be white. It is replaced and another draw is made. This ball also happens to be white. This is also replaced, and a simultaneous draw of two balls is made. What is the chance that both of them would be white.*

SOLUTION. The probability that 5, 4, 3, 2 or 1 ball is white is equal or it is $\frac{1}{5}$

(*a*) The probability that there are 5 white balls and 2 successive draws, after replacement would both give white balls

$$=\frac{1}{5}\times\frac{5}{5}\times\frac{5}{5}=\frac{25}{125}$$

(*b*) The probability that there are 4 white balls and 2 successive draws, after replacement would both give white balls

$$=\frac{1}{5}\times\frac{4}{5}\times\frac{4}{5}=\frac{16}{125}$$

(c) The probability that there are three white balls and 2 successive draws, after replacement would both give white balls

$$=\frac{1}{5}\times\frac{3}{5}\times\frac{3}{5}=\frac{9}{125}$$

(*d*) The probability that there are 2 white balls and 2 successive draws, after replacement would both give white balls

$$=\frac{1}{5}\times\frac{2}{5}\times\frac{2}{5}=\frac{4}{125}$$

(*e*) The probability that there is 1 white ball and 2 successive draws, after replacement, would both give white balls

$$=\frac{1}{5}\times\frac{1}{5}\times\frac{1}{5}=\frac{1}{125}$$

Thus the total of the probabilities of the first event

$$=\frac{25}{125}+\frac{16}{125}+\frac{9}{125}+\frac{4}{125}+\frac{1}{125}=\frac{55}{125}$$

The probability of event (*a*) and that of drawing two white balls simult neously

$$\frac{25}{125}\times\frac{^5C_2}{^5C_2}=\frac{150}{1250}$$

The probability of event (*b*) and that of drawing two white balls simultaneously

$$=\frac{16}{125}\times\frac{^4C_2}{^5C_2}=\frac{96}{1250}$$

The probability of event (*c*) and that of drawing two white balls simultaneously

$$=\frac{9}{125}\times\frac{^3C_2}{^5C_2}=\frac{27}{1250}$$

The probability of event (*d*) and that of drawing 2 white balls simultaneously

$$\frac{4}{125}\times\frac{^2C_2}{^5C_2}=\frac{4}{1250}$$

The total of these probabilities

$$=\frac{250}{1250}+\frac{96}{1250}+\frac{27}{1250}+\frac{4}{1250}=\frac{377}{1250}$$

∴ The probability of the events

$$=\frac{377}{1250}\div\frac{55}{125}$$

$$=\frac{377}{1250}\times\frac{125}{55}=\frac{377}{550}$$

PROBLEM 11.94. *A man has 5 rupee coins and one of them is known to have two heads. He takes out a coin at random and tosses it 5 times and it always falls head upward. What is the chance that it is the rupee with two heads ?*

SOLUTION. The probability that the rupee with two heads was selected and it fell heads all the five times

$$=\frac{1}{5}\times\frac{1}{1}\times\frac{1}{1}\times\frac{1}{1}\times\frac{1}{1}\times\frac{1}{1}=\frac{1}{5}$$

The probability that a rupee with one head was selected and it fell heads all the five times.

$$=\frac{4}{5}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{40}$$

∴ The chance that the rupee with 2 heads was selected

$$=\frac{1/5}{1/5+1/40}=\frac{1}{5}\times\frac{40}{9}=\frac{8}{9}$$

PROBLEM 11.95. *A speaks the truth in 75% cases and B in 80% of the cases. In what percentage of cases are they likely to contradict each other in stating the same fact ?*

SOLUTION. They will contradict each other only if one of them speaks the truth and the other speaks a lie.

The probability that A speaks the truth and B a lie

$$=\frac{75}{100}\times\frac{20}{100}=\frac{3}{20}$$

The probability that B speaks the truth and A a lie

$$=\frac{25}{100}\times\frac{80}{100}=\frac{4}{20}$$

∴ The total probability $=\frac{7}{20}$

∴ The percentage of cases in which they contradict each other

$$= \frac{7}{20} \times 100 = 35\%$$

PROBLEM 11.96. *A and B are two very weak students of statistics and their chances of solving a problem correctly are $\frac{1}{8}$ and $\frac{1}{12}$ respectively. They are given a question and obtain the same answer. If there probability of their making a common mistake is $\frac{1}{1001}$, find the chance that their answer was correct.*

SOLUTION. The probability that both A and B were right

$$= \frac{1}{8} \times \frac{1}{12} = \frac{1}{96}$$

The probability that both of them were wrong and committed a common

$$\text{mistake} = \frac{7}{8} \times \frac{11}{12} \times \frac{1}{1001} = \frac{77}{96096}$$

∴ The probability that both of them were right.

$$= \frac{1/96}{1/96 + 77/96096} = \frac{1}{96} \div \frac{1078}{96096} = \frac{1}{96} \times \frac{96096}{1078} = \frac{13}{14}$$

PROBLEM 11.97. *A bag contains 5 red and 3 black balls and a second one 4 red and 5 black balls. One of the bags is chosen at random and a draw of 2 balls is made from it. Find the chance that one is red and the other black.*

SOLUTION. The probability that the 1st bag was selected and a draw of 2 balls gives one red and one black ball.

$$= \frac{1}{2} \times \frac{{}^5C_1 \times {}^3C_1}{{}^8C_2} = \frac{15}{56}$$

The probability that the 2nd bag was selected and a draw of 2 balls gives one red and one black ball.

$$= \frac{1}{2} \times \frac{4C_1 \times {}^5C_1}{{}^9C_2} = \frac{5}{18}$$

∴ Since the events are mutually exclusive the probability of the event

$$= \frac{15}{56} + \frac{5}{18} = \frac{275}{504}$$

PROBLEM 11.98. *A and B play for a prize of Rs. 324. A is to throw a die first and is to win if he throws 6. If he fails B is to throw and is to win if he throws 6 or 5. If he fails, A is to throw again and to win if he throws 6, or 4 and so on. Find their respective expectations.*

SOLUTION.

Probability of A's winning in the first throw $= \frac{1}{6}$

" B's " " 2nd " $= \frac{5}{6} \times \frac{2}{6} = \frac{5}{18}$

" A's " " 3rd " $= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} = \frac{5}{18}$

" B's " " 4th " $= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{4}{6} = \frac{5}{27}$

" A's " " 5th " $= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{5}{6} = \frac{25}{324}$

" B's " 6th $= \frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{1}{6} \times \frac{6}{6} = \frac{5}{324}$

Total of A's chances of success $= \frac{1}{6} + \frac{5}{18} + \frac{25}{324} = \frac{169}{324}$

,, B's ,, ,, ,, $= \frac{5}{18} + \frac{5}{27} + \frac{5}{324} = \frac{155}{324}$

∴ Their respective expectations are

$A = \frac{169}{324} \times 324 = \text{Rs. } 169.$

$B = \frac{155}{324} \times 324 = \text{Rs. } 155.$

PROBLEM 11.99. *Eight mice are selected at random from a large numbe and then divided into two groups of four each—group A and group B. Each mouse i group A is given a dose 'a' of a certain poison which is expected to kill one in fou Each mouse in group B is given a dose 'b' of another poison which is expected kill one in two. Show that, nevertheless, there may be fewer deaths in group B tha in group A and find the probability of the happening*.

SOLUTION. There can be fewer deaths in group B in the followin ways :—

No. of deaths in Group B	No. of deaths in Group A
0	more than 0 (1, 2, 3 or 4)
1	more than 1 (2, 3 or 4)
2	more than 2 (3 or 4)
3	more than 3 (4)

To find out the probabilities of each of the above happenings we shall have to raise two binomials—one for group A, and the other for group B.

Group A

If 'p' stands for deaths and 'q' for survival,

$$(p+q)^3 = \left(\frac{1}{4} + \frac{3}{4}\right)^4$$

$$= \frac{1}{256} + \frac{12}{256} + \frac{54}{256} + \frac{108}{256} = \frac{175}{256}$$

Group B

$$(p+q)^3 = \left(\tfrac{1}{2} + \tfrac{1}{2}\right)^4$$

$$= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{15}{16}$$

The probability of the first event when there is no death in group B and more than zero, (1, 2, 3 or 4) deaths in group A,

$$= \left(\frac{1}{16} \times \frac{108}{256}\right) + \left(\frac{1}{16} \times \frac{54}{256}\right) + \left(\frac{1}{16} \times \frac{12}{256}\right) + \left(\frac{1}{16} \times \frac{1}{256}\right) = \frac{175}{4096}$$

The probability of the second event

$$= \left(\frac{4}{16} \times \frac{54}{256}\right) + \left(\frac{4}{16} \times \frac{12}{256}\right) + \left(\frac{4}{16} \times \frac{1}{256}\right) = \frac{268}{4096}$$

The probability of the third event

$$= \left(\frac{6}{16} + \frac{12}{256}\right) + \left(\frac{6}{16} \times \frac{1}{256}\right) = \frac{78}{4096}$$

The probability of the fourth event

$$= \left(\frac{4}{16} \times \frac{1}{256}\right) = \frac{4}{4096}$$

As the events are actually exclusive the probability of any of the events

$$= \frac{175}{4096} + \frac{268}{4096} + \frac{78}{4096} + \frac{4}{4096} = \frac{525}{4096}$$

REVISIONAL PROBLEMS—11

11.100. What is the chance that a leap year, selected at random, will contain 53 Mondays ? (*Ans* . $\frac{2}{7}$)

11.101. Two cards are drawn from a well-shuffled pack of 52 cards. Find the probability that they are both aces if the first card is (*a*) replaced (*b*) not replaced. [*Ans* . (*a*) $\frac{1}{169}$ (*b*) $\frac{1}{221}$]

11.102. A six faced die is so biased that it is twice as likely to show an even number as an odd number when thrown. It is thrown twice. What is the probability that the sum of the two numbers thrown is even. (*Ans* . $\frac{5}{9}$)

11.103. A can hit a target 3 times in 6 shots, B 2 times in 5 shots and C 3 times in 4 shots. They fire a volley. What is the probability that 2 shots hit ? (*Ans* : 9/20)

11.104. (*a*) What is the probability that a vowel selected at random in an English book is an *i* ?

(*b*) In a single throw with two dice, find the chances of throwing (*i*) 8 and (*ii*) 11.

(*c*) Four cards are drawn without replacement. What is the probability that they are all aces ? [*Ans* : $p=\frac{1}{26}$; (*b*) (*i*) $p=\frac{5}{36}$ (*ii*) $p=\frac{1}{18}$; (*c*) $p=\frac{1}{270725}$]

11.105. A problem in Statistics is given to three students A, B and C whose chances of solving it are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ respectively. What is the probability that the problem will be solved ? (*Ans* . $p=\frac{3}{4}$)

11.106. A, B and C in order toss a coin. The first one to throw a head wins. What are their respective chances to win. (*Ans* . $A=\frac{4}{7}$, $B=\frac{2}{7}$, $C=\frac{1}{7}$)

11.107. In a single throw with two dice, what is the probability of throwing (*a*) two aces and (*b*) 7 ? [*Ans* : (*a*) $\frac{1}{36}$ (*b*) $\frac{1}{6}$]

11.108. Five men in a company of 20 are graduates. If 3 men are picked out of 20 at random, what is the probability that they are all graduates ? What is the probability of at least one graduate ? (*Ans* . $p=1/4$, $p=\frac{137}{228}$)

11.109. (*a*) What is the probability of getting only one head in five tosses of a coin?

(*b*) The probability that a boy will get a scholarship is 0.90 and that a girl will get is 0.80. What is the probability that at least one of them will get the scholarship ? [*Ans* . (*a*) $p=\frac{5}{32}$ (*b*) $p=0.98$]

10.110. An urn contains 4 red and 3 blue balls. Two drawings of 2 balls are made. Find the chance that the first drawing gives 2 red balls, and the second drawing gives 2 blue balls, if the balls are returned to the urn after the first draw. (*Ans* : $p=\frac{2}{49}$)

11.111. Suppose two six-faced dice are thrown 10 times. What is the probability of getting double six in at least one of the throws ? [*Hint* . $p=\{1-(\frac{35}{36})^{10}\}$]

11.112. If a single draw is made from a pack of 52 cards, what is the probability of securing either an ace of spades or jack of clubs ? (*Ans* : $p=\frac{1}{26}$)

11.113. A and B decide to meet between 5 and 6 p.m. but that each should wait no longer than 10 minutes for the other. Determine the probability that they meet. (*Ans* . $p=\frac{11}{36}$)

11.114. Two cards are randomly drawn from a pack of 52 cards and thrown away. What is the probability of drawing an ace in a single draw from the remaining 50 cards ? (*Ans* . $p=\frac{1}{20825}$)

11.115. A bag contains 5 white and 4 black balls. A ball is drawn from this bag and replaced, and then a second draw of a ball is made. What is the probability that one of these is red and the other is black ? [*Ans* . $p=\frac{275}{1008}$]

Probability (Contd.) 12

In this chapter are some comparatively difficult questions of probability. The treatment of the subject in this chapter is mathematical. As such it can be omitted by students who need to have an elementary knowledge of the subject only. Two or three questions of the previous chapter have been repeated to illustrate the mathematical treatment.

MATHEMATICAL TREATMENT OF PROBABILITY

By Direct Enumeration of Cases :

PROBLEM 12.1. *What is the probability of getting 9 points with 3 dice ?*

SOLUTION. If 3 dice are rolled, the number of all possible cases $=6\times6\times6=6^3$

Let the first die show up a_1 ; the second a_2 ; and the third a_3 where a_1, a_2 $a_3=1, 2, 3, 4, 5, 6$.

If the sum of the points is to be 9, a_1, a_2, a_3 should be such that a_1+a_2 $+a_3=9$, where $a_1, a_2, a_3=1, 2, 3, 4, 5, 6$.

The number of all favourable cases is equal to the number of solutions of this equation.

To find out the number of solutions of this equation, consider the polynomial :

$$x+x^2+x^3+x^4+x^5+x^6$$

Multiplying it by itself, we get,

$$(x^{1+1}+x^{1+2}+\ldots\ldots+x^{1+6})+x^{2+1}+\ldots\ldots+x^{2+6})\pm\ldots.+(x^{6+1}+x^{6+2}+\ldots.+x^{6+6})$$

It may be seen that exponents of x are a_1+a_2 where $a_1, a_2=1, 2, 3, 4,$ 5, 6.

Similarly, the exponents of

$$(x+x^2+x^3+x^4+x^5+x^6)^3$$

will be $a_1+a_2+a_3$, where $a_1, a_2, a_3=1, 2, 3, 4, 5, 6$.

Now, we have to find the number of all those cases for which

$$a_1+a_2+a_3=9$$

and this obviously is the coefficient of x^3 in the expansion of

$$(x+x^2+x^3+x^4+x^5+x^6)^3$$

Now, $x+x^2+x^3+x^4+x^5+x^6=\dfrac{x(1-x^6)}{1-x}$(Sum of G.P.)

$$\therefore\ (x+x^2+x^3+x^4+x^5+x^6)^3=\frac{x^3(1-x^6)^3}{(1-x)^3}$$

By the binomial theorem :

$$x^3(1-x^6)^3=x^3[1-3x^6+3x^{12}-x^{18}]$$

and $(1-x)^{-3}=1+3x^2+6x^2+10x^3+15x^4+21x^5+28x^6\ldots.$

$$\therefore x^3(1-x^6)^3/(1-x)^3=[x^3-3x^9+\ldots..\times[1+3x+\ldots+28x^6+\ldots]$$

Multiplying the expressions on *r. b. s.* and collecting the terms x^2 we find that the coefficient of

$$x^3 = 28 - 3 = 25$$

∴ The number of solutions of the equation = No. of all favourable cases = 25.

∴ The required probability $= \frac{25}{6^3} = \frac{25}{216}$

[*N. B.* This technique is meant for those cases where the number of dice is large. Here it is used for illustration. Consider the next problem, where the enumeration of favourable cases, without using this device, is very difficult.]

PROBLEM 12.2. *What is the probability of getting 20 points with 6 dice ?*

SOLUTION. If 6 dice are rolled, the number of all possible cases $= 6\times6\times6\times6\times6\times6 = 6^6$.

Let the first die show up a_1 ; the second, a_2 ; the third, a_3; the fourth, a_4 ; the fifth, a_5 ; and the sixth, a_6—where $a_1, a_2, a_3, a_4, a_5, a_6 = 1, 2, 3, 4, 5, 6$.

Now $a_1, a_2, a_3, a_4, a_5, a_6$ should be such that

$a_1 + a_2 + a_3 + a_4 + a_5 + a_6 = 20$.

The number of all favourable cases is equal to the number of solutions of this equation. To find out this number consider the polynomial :

$$x + x^2 + x^3 + x^4 + x^5 + x^6$$

The number of all favourable cases will be equal to the coefficient of x^{20} in the expansion of

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^6$$

Now, $x + x^2 + x^3 + x^4 + x^5 + x^6 = \frac{x(1-x^6)}{1-x}$ Sum of G. P.

$$\therefore (x + x^2 + x^3 + x^4 + x^5 + x^6)^6 = \frac{x^6(1-x^6)^6}{(1-x)^6}$$

By the binomial theorem :

$$x^6(1-x^6)^6 = x^6[1 - {}^6C_1x^6 + {}^6C_2x^{12} - {}^6C_3x^{18} + {}^6C_4x^{24} - \ldots\ldots + x^{36}]$$
$$= x^6 - 6x^{12} + 15x^{18} - 20x^{24} + \ldots\ldots + x^{42}$$

and $\frac{1}{(1-x)^6} = (1-x)^6 = 1 + 6x + \frac{6.7}{2.1}x^9 + \ldots\ldots\ldots$

$$+ \frac{6\times7\times8\times9\times10\times11\times12\times13}{8\times7\times6\times5\times4\times3\times2\times1}x^3 + \ldots\ldots$$

$$+ \frac{6\times7\times8\times9\times10\times11\times12\times13\times14\times15\times16\times17\times18\times19}{14\ !}x^{14} + \ldots\ldots$$

$$= 1 + 6x + 21x^2 + \ldots\ldots + 1287x^8 + \ldots\ldots + 11628x^{14} +$$

$$\therefore \left[\frac{x(1-x^6)}{1-x}\right] = [x^6 - 6x^{12} + 15x^{18} - 20x^{24} + \ldots] \times [1 + 6x + 21x^2 + \ldots\ldots$$
$$+ 1287x^8 + \ldots\ldots + 11628x^{14} + \ldots\ldots]$$

Multiplying the expressions on *r, b, s.* and collecting the terms x^{20} we find :

$$11628x^{20} - 6.1287x^{20} + 15.21x^{20}$$

∴ Coefficient of $x^{20} = (4628 - 7722 + 315)$

$$= 11943 - 7722 = 4221$$

∴ The number of solutions = The number of all favourable cases = 4221

∴ The required probability $= \frac{4221}{6^6} = \frac{4221}{46656} = \frac{469}{5184}$

PROBLEM 12.3. *Twelve balls are distributed at random among three boxes. What is the probability that the first box contains three balls ?*

SOLUTION. To distinguish one ball from other, let us number the 1, 2, 3, ..., 11, 12.

Now Ball 1 may fall in any of the 3 boxes. This gives 3 cases.

Ball 2 may fall in any of the 3 boxes. Therefore, for two balls there wi be $3 \times 3 = 3^2$ cases and so on. For twelve balls there will be 3^{12} possible cases.

Now, suppose the first box contains 3 balls. The remaining 9 balls ma be distributed in 2 boxes in any of the ways.

The number of ways in which 9 balls may be distributed in 2 boxes$=2^9$.

Next, there are 12 balls and 3 balls are in a box. Out of 12 balls 3 bal can be selected in $^{12}C_3$ ways.

$\therefore$ Number of all favourable cases$= {}^{12}C_3 \times 2^9$.

$$= \frac{12 \times 11 \times 10}{3 \times 2 \times 1} \times 2^9$$

$\therefore$ The required probability $= \dfrac{12 \times 11 \times 10 \times 2^9}{3 \times 2 \times 1 \times 3^{12}}$

$$= \frac{112640}{531441}$$

PROBLEM 12.4. *A certain number (200) of identical balls is distribute among 20 compartments. What is the most probable number of balls in a specif compartment?*

SOLUTION. Let the most probable number of balls be x.

No. of all possible cases $= 20^{200}$

No. of all favourable cases$= {}^{200}C_x \times (20-1)^{200-x}$

$\therefore$ The required probability $= \dfrac{^{200}C_x (20-1)^{200-x}}{20^{200}}$

$$= \frac{^{200}C_x \times 19^{200-x}}{20^{200}}$$

This will be the greatest if $^{200}C_x \times 19^{200-x}$ is the greatest.

i.e. if, $200C_x \times 19^{200-x} \geqslant {}^{200}C_{x-1} \times 19^{200-x+1}$

$200C_x \times 19^{200-x} \geqslant {}^{200}C_{x+1} \times 19^{200-1}$

(for) $200C_x \times 19^{200-x}$ is xth term in the expansion of $(19+11)^{200}$

of if $\dfrac{200!}{x!(200-x)!} \times 19^{200-x} \geqslant \dfrac{200\,!}{(x-1)!\,(200-x+1)!} \times 19^{201-x}$

$\dfrac{200\,!}{x!(200-x)!} \times 19^{200-x} \geqslant \dfrac{200!}{(x+1)!\,(200-x+1)\,!} \times 19^{199-x}$

or, if $\dfrac{1}{x} \geqslant \dfrac{1}{201-x} \times 19$ or $201-x \geqslant 19x$, or $201 \geqslant 20x$

$\dfrac{1}{200-x} \times 19 \dfrac{1}{x-1}$ or $19x+19 \geqslant 200-x$, or $20x \geqslant 181$

i.e., $181 \geqslant 20x \geqslant 201$

$\therefore x = 10$.

The required probability $P = \dfrac{^{200}C_{10} \times 19^{190}}{20^{200}}$

$$= \frac{200\,!}{190\,!\,10\,!} \times \frac{19^{190}}{20^{200}}$$

Anti-log [Log $P = 259.3233 - 260.2000$] $= \bar{1}.1233$

$\therefore P = 0.1328$

PROBLEM 12.5. *An urn contains 12 white balls and 8 black balls. If 8 balls are drawn from this urn, find the probability that among them there will be exactly 5 white balls and 3 black balls.*

SOLUTION. The total number of ways to get 8 balls out of 20 balls is ${}^{20}C_8$. This is the number of all possible and equally likely cases.

5 white balls can be drawn out of 12 white balls in the urn in ${}^{12}C_5$ ways, and 3 black balls can be drawn out of 8 black balls in ${}^{8}C_3$ ways

Now every group of 5 white balls combines with every possible group of 3 black balls.

$\therefore$ The no. of ways to form all groups containing 5 white balls and 3 black balls $= {}^{12}C_5 \times {}^{8}C_3$.

This is the number of all favourable cases.

$\therefore$ The required probability $= \dfrac{{}^{12}C_5 \times {}^{8}C_3}{{}^{20}C_8}$

$$= \frac{12\times 11\times 10\times 9\times 8}{5\times 4\times 3\times 2\times 1} \times \frac{8\times 7\times 6}{3\times 2\times 1} \times \frac{1\times 2\times 3\times 4\times 5\times 6\times 7\times 8}{20\times 19\times 18\times 17\times 16\times 15\times 14\times 13}$$

$$= \frac{7392}{20995}$$

PROBLEM 12.6. *What is the probability of obtaining a sum of 10 in a single throw of five dice ?*

SOLUTION. A sum of 10 can be obtained in any of the following ways:
Combinations giving a sum of 10 :

(1) 1, 1, 1, 1, 6
(2) 1, 1, 1, 2, 5
(3) 1, 1, 1, 3, 4
(4) 1, 1, 2, 3, 3
(5) 1, 1, 2, 2, 4
(6) 1, 2, 2, 2, 3
(7) 2, 2, 2, 2, 2

Now the first combination has got $\dfrac{5\,!}{4\,!} = 5$ Permutations

,, second ,, ,, ,, $\dfrac{5\,!}{3\,!} = 20$,,

,, third ,, ,, ,, $\dfrac{5\,!}{8\,!} = 20$,,

,, fourth ,, ,, ,, $\dfrac{5\,!}{2!\;2!} = 30$,,

the fifth combination has got $\dfrac{5\,!}{2\,!\,2\,!} = 30$ Permutations

,, sixth ,, ,, ,, $\dfrac{5\,!}{3\,!} = 20$,,

,, seventh ,, ,, ,, $\dfrac{5\,!}{5\,!} = 1$,,

$\therefore$Total number of favourable cases $= 126$

Number of all possible cases $= (6)^5 = 7776$

$\therefore$the required probability $= \dfrac{126}{7776} = \dfrac{21}{1296} = \dfrac{7}{432}$

PROBLEM 12.7. *A triangular prism, rectangular faces marked 1, 2, 3 respectively ; a tetrahedral die, 1, 2, 3, 4 respectively, and an ordinary cubical die are tossed. The spots on the under faces of the prism and tetrahedral and on the upper face of the cubical die are counted. What is the probability of a sum of 7 ?*

SOLUTION. A sum of 7 could be made up of any of the following combinations :

(1)	1,	1,	5
(2)	1,	2,	4
(3)	3.	3,	1
(4)	2,	2,	3

The first combination can occur in 1 way only
,, second ,, ,, ,, 4 ways ,,
,, third ,, ,, ,, 3 ways ,,
,, fourth ,, ,, ,, 3 ways ,,

∴No. of all favourable cases $= 11$
No. of all possible cases $= 3 \times 4 \times 6 = 72$

∴The required probability $= \dfrac{11}{72}$

PROBLEM 12.8. *20 letters, to each of which corresponds an envelope, are placed in the envelopes at random. What is the probability that letter one is placed in the right envelope ?*

SOLUTION. Let the number of ways in which all letters can go wrong be $\mu 20$. Now consider any particular letter. If this occupies another's envelope and *vice versa*, this can happen in $(20-1) = 19$ ways. The remaining $(20-2)$ letters can go wrong in $\mu 20-2 = \mu 18$ ways. But if the letter occupies another's place, this can happen in $20-1 = 19$ ways and if not *vice versa*, there are $\mu_{20}-1 = \mu_{19}$ way in which others can go wrong.

$$\therefore \mu 20 = (20-1)\ (\mu 19 + \mu 18)$$
$$= 19(\mu 19 + \mu 18)$$

This may be written as

$$\mu 20 = 20\mu 19 - \mu 19 + 20\mu 18 - \mu 18$$

or $\mu 20 - 20\mu 19 = -[\mu 19 - (20-)\mu 18]$

Let $\mu 20 = \mu 20 - 20\mu 19$

Then $\mu 20 = -\mu 19$

$$= (-1)^{20-2}\mu_2$$

Thus $\mu 20 - 20\mu 19 = (-1)^{20-2}(\mu 2 - \mu 1)$

Now $\mu 1$, *i.e.* the number of ways in which 1 letter can go wrong is equal to zero ; and $\mu 2$, *i.e.* the number of ways in which two letters can go wrong, is 1.

$$\therefore \mu 20 - 20\mu 19 = (-1)^{20-2} = (-1)^{20-2}(-1)^2 = (-1)^{20}$$
$$\mu 20 - 20\mu 19 = (-1)^{20}$$

or $\mu 20 = 1 + 20\mu 19$

$$1 + 20[(-1)^{19} + 19\mu 18]$$
$$= 1 - 20 + 20.19[(-1)^{18} + 18\mu 17]$$
$$= 1 - 20 + 20.19 + 20.19.18[(-1)^{17} + 17\mu 16]$$
$$= \ldots\ldots\ldots\ldots\ldots\ldots$$
$$= 1 - 20 + (20 \times 19) - (20 \times 19 \times 18) + \ldots + (20 \times 19 \times 18).5\ [(-1)^4 + 4\mu 3]$$
$$= 1 - 20 + (20 \times 19) - (20 \times 19 \times 18) + \ldots + (20 \times 19 \times \ldots\ldots 5) + 20 \times 19 \times 18 \ldots 5 \times 4[(-1)^3 + 3\mu 2]$$
$$= 1 - 20 + (20 \times 19) - (20 \times 19 \times 18) + \ldots + (20 \times 19 \times \ldots 5) - (20 \times 19 \times 18 \times \ldots 5 \times 4) + 20 \times 19 \ldots\ldots \times 4 \times 3[(4-1)^2 + 2\mu 1]$$
$$= 1 - 20 + (20 \times 19) - (20 \times 19 \times 18) + \ldots\ldots + (20 \times 19 \times \ldots 4 \times 3)$$

$$=\frac{20\,!}{20\,!}-\frac{20\,!}{19\,!}+\frac{20\,!}{18\,!}-\frac{20\,!}{17\,!}+\ldots\ldots+\frac{20\,!}{2\,!}$$

$$=20\,!\left[\frac{1}{2\,!}-\frac{1}{3\,!}+\frac{1}{4\,!}-\ldots\ldots+\frac{1}{18\,!}-\frac{1}{19\,!}+\frac{1}{20\,!}\right]$$

Now the total number of possible ways $=20\,!$

$$\therefore \text{The required probability}=\frac{20\,!\left[\frac{1}{2\,!}-\frac{1}{3\,!}+\ldots\ldots+\frac{1}{10!}\right]}{20\,!}$$

$$=\frac{1}{2\,!}-\frac{1}{3\,!}+\frac{1}{4\,!}-\ldots\frac{1}{19\,!}+\frac{1}{20\,!}$$

i. e. the first twenty terms of e^{-1}.

PROBLEM 12.9. *Twelve persons, amongst whom are A & B, sit down at random at a round table. What is the probability that there are r persons—r=0, 1, 2, etc.—between A and B ?*

SOLUTION. A and B can occupy ${}^{12}P_2=12\times 11=132$ positions in all. This is the number of all possible cases.

Seats occupied

0 (1,2), (2,3), (3,4), (4,5), (5,6), (6, 7), (7,8), (8,9), (9,10), (10,11), (11,12), (12,1)

1 (1,3), (2,4), (3,5),……(10, 12), (11,1), (12,2)

2 (1,4), (2,5), (3,6),……(10,1), (11,2), (12,3)

3 (1,5), (2,6), (3,7),……(10,2), (11,3), (12,4)

4 (1,6), (2,7), (3,8),……(10,3), (11,4) (12,5)

5 (1,7), (2,8) (3,9),……(10,4), (11,5), (12,6)

when $r=0$—A and B can interchange seats, ∴ No, of all favourable cases$=2\times 12$

" $r=1$—A and B can interchange seats, ∴ No. of all favourable cases$=2\times 12$

" $r=2$—A and B can interchange seats, ∴ No. of all favourable cases$=2\times 12$

" $r=3$—A and B can interchange seats, ∴ No. of all favourable cases$=2\times 12$

" $r=4$—A and B can interchange seats, ∴ No, of all favourable cases$=2\times 12$

" $r=5$—The positions that they take after interchanging seats is already counted, and therefore no. of all favourable cases$=12$

For r 75, we observe that these cases are the same as those for $r<5$, (see the adjoining diagram).

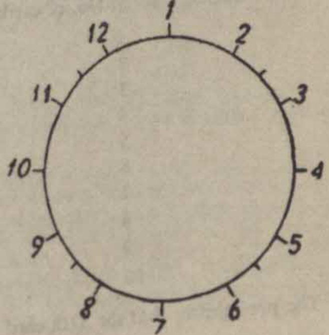
Thus when $r=0$, there are $r=10$ persons between A and B

Thus when $r=1$, there are $r=9$ persons between A and B

Thus when $r=4$, there are $r=6$ persons between A and B

The required probability

$$\text{for } r=0,=\frac{12\times 1}{11\times 12}=\frac{1}{11}$$



$$\text{for } r=1, = \frac{2\times 12}{11\times 12}=\frac{2}{11} \quad \text{for } r=2, = \frac{2\times 12}{11\times 12}=\frac{2}{11}$$

$$\text{for } r\ 3, = \frac{2\times 12}{11\times 12}=\frac{2}{11}$$

$$\text{for } r\ 4, = \frac{2\times 12}{11\times 12}=\frac{2}{11}$$

$$\text{for } r=5, = \frac{2\times 12}{11\times 12}=\frac{2}{11}$$

PROBLEM 12.10. *A and B stand in line, at random, with 10 other perso What is the probability that there will be r persons—r=0, 1, 2 etc. betwe them ?*

SOLUTION. A and B can take $12P2 = 12\times 11$ positions. This is the numb of all possible cases.

Next, we find out the number of positions for different values of r. It ma be noted that A and B can interchange positions.

Position taken	No. of Favourable cases	Probability
0. (1,2),(2,3),(3,4),(5,6),(6,7),(7,8), (8,9),(9,10),(10,11),(11,12)	2×11	$\frac{22}{132}=\frac{1}{6}$
1. (1,3), (2,4)....(9,11), (10,12)	2×10	$\frac{20}{132}=\frac{5}{33}$
2. (1,4), (2,5).....(9,12)	2×9	$\frac{18}{132}=\frac{3}{22}$
3. (1,5), (2,6) ..(8,12)	2×8	$\frac{16}{132}=\frac{4}{33}$
4. (1,6), (2,7)....(7,12)	2×7	$\frac{14}{132}=\frac{7}{66}$
5. (1,7), (2,8)....(6,12)	2×6	$\frac{12}{132}=\frac{1}{11}$
6. (1,8), (2,9)....(5,12)	2×5	$\frac{10}{132}=\frac{5}{66}$
7. (1,9), (2,10), (3,11), (4,12)	2×4	$\frac{8}{132}=\frac{2}{33}$
8. (1,10),(2,11), (3,12)	2×3	$\frac{6}{132}=\frac{1}{22}$
9. (1,11), (2,12)	2×2	$\frac{4}{132}=\frac{1}{33}$
10. (1,12)	2×1	$\frac{2}{132}=\frac{1}{66}$

Total and Compound Probability

PROBLEM 12.11. *Cards are drawn one by one from a full deck. What is the probability that 10 cards will precede the first ace ?*

SOLUTION.

Serial No. of cards.	The probability that it is not an ace
1	48/52
2	47/51
3	46/50
4	45/49
5	44/48
6	43/47
7	42/46
8	41/45
9	40/44
10	39/43

The probability that the 11th card is an ace $=\frac{4}{42}$

$$\therefore \text{the required probability} = \frac{48}{52} \times \frac{47}{51} \times \frac{46}{50} \times \frac{45}{49} \times \frac{44}{48} \times \frac{43}{47} \times \frac{42}{46} \times \frac{41}{45} \times \frac{40}{44} \times \frac{39}{43} \times \frac{4}{42}$$

$$= \frac{41 \times 40 \times 39 \times 4}{52 \times 51 \times 50 \times 49}$$

$$= \frac{164}{4165}$$

PROBLEM 12.12. *Urn 1 contains 10 white and 3 black balls ; urn 2 contains 3 white and 5 black balls. Two balls are transferred from urn 1 and placed in urn 2 and then one ball is taken from the latter. What is the probability that it is a white ball ?*

SOLUTION. The probability of taking out a white ball from $\text{urn}_1 = \frac{10}{13}$

" " " " " black " " $= \frac{3}{13}$

Now, two balls are transferred from urn_1 to urn_2

The probability that both of them are white $= \frac{10}{13} \cdot \frac{9}{12}$

" " " " " " black $= \frac{3}{13} \cdot \frac{2}{12}$

Now, one white and one black ball can be taken out in two ways : first white and second black, first black and second white.

$\therefore$ The probability that one of them is white and the other black

$$= \left(\frac{10}{13} \times \frac{3}{12} \right) + \left(\frac{3}{13} \times \frac{10}{12} \right) = \frac{5}{13}$$

corresponding to these three cases, in which there are respectively 5 white balls, 5 black balls, 3 white balls, 7 black balls and 4 white balls and 6 black balls in urn_2, we shall have three probabilities of drawing a white ball from urn_2.

The probability of drawing a white ball in the

" " " " " " first case $= \frac{10}{13} \times \frac{9}{12} \times \frac{5}{10}$

" " " " " " second case $= \frac{3}{13} \times \frac{2}{12} \times \frac{3}{10}$

" " " " " " third case $= 2 \times \frac{3}{12} \times \frac{10}{13} \times \frac{4}{10}$

$\therefore$The probability that the ball taken from urn_2 is white

$$= \left(\frac{10}{13} \times \frac{9}{12} \times \frac{5}{10} \right) + \left(\frac{3 \times 2 \times 3}{13 \times 12 \times 10} \right) + \left(\frac{2 \times 3 \times 4 \times 10}{13 \times 12 \times 10} \right)$$

$$= \frac{450}{1560} + \frac{18}{1560} + \frac{240}{1560} = \frac{59}{130}$$

PROBLEM 12.13. *A, B and C in order toss a coin. The first one to throw a head wins. What are their respective chances of winning ? Assume that the game may continue indefinitely.*

SOLUTION. The probability that A throws a head in the first toss $= \frac{1}{2}$. If he throws a head, he wins. But if he does not throw a head in the the first toss, B tosses the coin. Therefore, B's probability of winning = (A's probability of not throwing head (×) B's probability of throwing a head) $= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. If he throws a head he wins. But if he does not throw a head C's turn comes. C's probability of winning = B's probability of not throwing a head) × (C's probability of not throwing a head) $\frac{1}{4} \times \frac{1}{2}, = \frac{1}{8}$. If C fails A's turn comes and so on. Their respective probabilities are given below:—

A	B	C
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$
$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$
$\frac{1}{128}$	$\frac{1}{256}$	$\frac{1}{512}$
...	...	...

$\therefore$ A's chances of winning $= \left(1+\frac{1}{8}+\frac{1}{64}+\ldots.\right)$

$= \frac{1}{2} \times \frac{1}{1-\frac{1}{8}}$ Sum of G. P.

$= \frac{4}{7}$

$\therefore$ B's chances of winning $= \frac{1}{4}\left(1+\frac{1}{8}+\frac{1}{64}+\ldots\ldots\right)$

$= \frac{1}{4} \times \frac{1}{1-\frac{1}{8}}$

$= \frac{2}{7}$

$\therefore$ C's chances of winning $= \frac{1}{8}\left(1+\frac{1}{8}+\frac{1}{64}+\ldots\ldots\right)$

$= \frac{1}{8} \times \frac{1}{1-\frac{1}{8}}$

$= \frac{1}{7}$

PROBLEM 12.14. *A bag contains 20 red balls and 15 white balls. If th are drawn one by one, find out the probability of drawing first a red, then a whi and so on alternately until only red balls are left.*

SOLUTION. The probability of drawing

1st red ball $= \frac{20}{20+15}$

1st white ball after red ball $= \frac{20}{20+15} \times \frac{15}{19+15}$

2nd red ball after 1 red and 1 white ball

$$= \frac{20}{20+15} \times \frac{15}{19+15} \times \frac{19}{19+14}$$

2nd white ball after 1 red, 1 white and 1 red ball

$$= \frac{20}{20+15} \times \frac{15}{19+15} \times \frac{19}{19+14} \times \frac{14}{18+14}$$

14th white ball after alternately drawing 1 red and 1 white.

$$\text{ball} = \frac{20}{20+15} \times \frac{15}{19+15} \times \frac{19}{19+14} \times \frac{14}{18+14} \times \cdot \times \frac{2}{6+2}$$

15th white ball after 15th red ball

$$= \frac{20}{20+15} \times \frac{15}{19+15} \times \frac{19}{19+14} \times \frac{14}{18+14} \times \cdots \times \frac{2}{6+2} \times \frac{2}{6+1}$$

15th white ball after 15th red ball

$$= \frac{20}{20+15} \times \frac{15}{19+15} \times \frac{19}{19+14} \times \frac{14}{18+14} \times \cdot\cdot \times \frac{2}{6+2} \times \frac{6}{6+1} \times \frac{1}{5+1}$$

Now that all the 15 white balls have been drawn only red balls are left i the bag.

The required probability

$$= \frac{20}{20+15} \times \frac{15}{19+15} \times \cdots \cdot \times \frac{6}{6+1} \times \frac{1}{5+1}$$

$$= \frac{(20 \times 19 \times 18 \times \ldots. \times 6)\,(15 \times 14 \times \ldots. \times 2.1)}{35 \times 34 \times \ldots. \times 6}$$

$$= \frac{20!\ \ 15!}{35!}$$

PROBLEM 12.15. *Two players A and B toss two dice, A starting the game. The game is won by A if he casts 6 points before B casts 7 points; and it is won by B if he casts 7 points before A casts 6 points, (i) what are the probabilities for A and B to win the game if they agree to cast the dice 20 times ? (ii) what is the probability of a tie ? and (iii) what are the probabilities for A and B to win the game if the game is continued indefinitely.*

SOLUTION. If two dice are cast the number of all possible cases $= 6\times 6 = 36$.

Cases favourable to 1st die	6 points 2nd die	Cases favourable to 1st die	7 points 2nd die
0	6	1	6
1	5	2	5
2	4	3	4
3	3	4	3
4	2	5	2
5	1	6	1
5 cases			6 cases

$\therefore$ Probability of casting 6 points with two dice$=\frac{6}{36}$

" " not " 6 " " " $=1-\frac{5}{36}=\frac{31}{36}$

" " casting " 7 " " " $=\frac{36}{36}$

" " not " 7 " " " $=1-\frac{6}{36}=\frac{30}{36}$

The probabilities of A's and B's winning are given below. It should be remembered that one will get chance of tossing the dice only when the other has failed.

A's Probability	B's Probability
$\frac{5}{36}$	$\frac{31}{36}\times\frac{6}{36}$
$(\frac{31}{36}\times\frac{30}{36})\times\frac{5}{36}$	$(\frac{31}{36}\times\frac{6}{36})\times\frac{31}{36}\times\frac{6}{36}$
$(\frac{31}{31}\times\frac{30}{36})^2\times\frac{5}{36}$	$(\frac{31}{36}\times\frac{30}{36})\times\frac{31}{36}\times\frac{6}{36}$
…………………	…………………
(19th trial)	(20th trial)
$(\frac{31}{36}\times\frac{30}{36})^9\times\frac{5}{36}$	$(\frac{31}{36}\times\frac{30}{36})^9\times\frac{31}{36}\times\frac{5}{36}$
…………………	…………………

(*i*) A's probability of winning in 20 tosses

$$=\frac{5}{36}\left[1+\frac{30\times 31}{36^2}+\left(\frac{30\times 31}{36^2}\right)^2+\ldots+\left(\frac{33\times 31}{36^2}\right)^9\right].$$

$$=\frac{5}{36}\left\{\frac{1-\left(\frac{30\times 31}{36^2}\right)^{10}}{1-\frac{30\times 31}{36^2}}\right\} \qquad \text{—Sum of G. P.}$$

$$=\frac{5}{36}\left[\frac{216}{61}\left(1-\left(\frac{155}{216}\right)^{10}\right)\right]$$

$$=^{30}\left[1-\left(\frac{155}{216}\right)^{10}\right]$$

B's probability of winning in 20 tosses

$$=\frac{6}{36}\times\frac{31}{36}\left[1+\frac{30\times 31}{36^2}+\left(\frac{30+31}{36^2}\right)^2+\ldots+\left(\frac{30\times 31}{36^2}\right)^9\right.$$

$$=\frac{6}{36}\times\frac{31}{36}\left[\frac{216}{61}\left(1-\left(\frac{155}{246}\right)^{10}\right)\right]$$

$$=\frac{31}{61}\left[1-\left(\frac{155}{116}\right)^{10}\right]$$

(*ii*) In twenty tosses either A will win or B will win or there will be a
These are the three mutually exclusive and exhaustive cases.

∴ 1=Probability that A wins+Probability that B wins+Probability
there is a tie.

∴ The probability of a tie

$$=1-\left[\frac{30}{61}\left\{1-\left(\frac{155}{216}\right)^{10}\right\}+\frac{31}{61}\left\{1-\left(\frac{155}{216}\right)^{10}\right\}\right]$$

$$=1-\left[1-\left(\frac{155}{216}\right)^{10}\right]=\left(\frac{155}{216}\right)^{10}$$

(*iii*) If the game continues indefinitely :

$$\text{A's Probability}=\frac{5}{36}\left[1+\left(\frac{30\times31}{36^2}\right)+\left(\frac{30\times32}{36^2}\right)+\ldots\ldots\right]$$

$$=\frac{5}{36}\times\frac{1}{1-\frac{155}{216}}=\frac{5}{36}\times\frac{216}{61}=\frac{30}{61}$$

$$\text{B's Probability}=\frac{6}{36}\times\frac{31}{36}\left[1+\frac{30\times31}{36^2}+\left(\frac{30\times31}{36^2}\right)^2+\ldots\right]$$

$$=\frac{6}{36}\times\frac{31}{36}\times\frac{1}{1-\frac{155}{216}}=\frac{6}{36}\times\frac{31}{36}\times\frac{216}{61}=\frac{31}{61}$$

PROBLEM 12.16. *Peter and Paul play a game with two dice. Pet plays first by throwing the dice together. If the total number of points is a pri number other than 2 he wins outright; if it is even he throws again under the sa conditions, in other cases the throw passes to Paul, who throws under the same co ditions. What is the probability of Peter's winning ?*

SOLUTION. First let us see the number of ways in which the thro occur.

First dice	Second dice
1	1
2	2
3	3
4	4
5	5
6	6

It is obvious that the possible throws with two dice are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

Total points :	2	3	4	5	6	7	8	9	10	11	12	Tota
No. of ways :	1	2	3	4	5	6	5	4	3	2	1	36

Probability of throwing a prime (*i.e.* 2, 3, 5, 7 and 11) other than a

$$=\frac{2}{36}+\frac{4}{36}+\frac{6}{36}+\frac{2}{36}=\frac{14}{36}$$

Probability of throwing an even number

$$=\frac{1}{36}+\frac{3}{36}+\frac{5}{36}+\frac{5}{36}+\frac{3}{36}+\frac{1}{36}=\frac{18}{36}$$

Probability of throwing neither$=\frac{4}{36}$

These three cases are mutually exclusive. Let the probability of Peter winning be P. If he throws a prime other than 2, he wins outright. The pr bability of this is $=\frac{14}{36}$

If he throws an even number, he throws again.

The probability that he throws an even number$=\frac{18}{36}$

and the probability that he wins$=\frac{18}{36}$P.

If he throws neither, the probability of which is $\frac{4}{36}$, the throw passes to Paul and his chance of winning is P, so that Peter's chance of winning is $\frac{4}{36}(1-P)$.

$$\therefore P=\frac{14}{36}+\frac{18P}{36}+\frac{4}{36}(1-P)$$

$$\text{or } P-\frac{18P}{36}+\frac{4P}{36}=\frac{14}{36}+\frac{4}{36}$$

$$\frac{22P}{36}=\frac{18}{36}$$

$$\text{or } P=\frac{18}{22}=\frac{9}{11}$$

Peter's chance of winning $=\frac{9}{11}$

PROBLEM 12.17. *If a person playing a certain game can win Re. 1 with the probability $\frac{1}{3}$, and lose 25 np. with the probability $\frac{2}{3}$, what is the probability of winning at least Rs. 3 in 20 games ?*

SOLUTION. Let x be the number of times the game is won. Therefore the game is lost $20-x$ times.

The total gain $=1.\,x-\frac{1}{4}(20-x)=(\frac{5}{4}x-5)$ rupees.

Now Rs. $(\frac{5}{4}x-5)$ should not be less than Rs. 3

or $\frac{5}{4}x-5\geqslant 3$

or $5x\geqslant 32$

or $x\geqslant 6\frac{2}{5}$

Since x must be an integer, therefore in $\geqslant 7$.

That is in 20 trials an event with probability $\frac{1}{3}$ must happen at least 7 times.

The probability for that is : $\displaystyle\sum_{x=7}^{20}\frac{20\,!}{x\,!\,(20-x)}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{20-x}$

$$=1-\sum_{x=0}^{6}\frac{20\,!}{x\,!\,(20-x)\,!}\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{20-x}$$

$$=1-\Big[(\tfrac{2}{3})^{20}+\frac{20\,!}{1\,!\,19\,!}\tfrac{1}{3}(\tfrac{2}{3})^{19}+\frac{20\,!}{2\,!\,18\,!}(\tfrac{1}{3})^{2}(\tfrac{2}{3})^{18}+\frac{20\,!}{3\,!\,17\,!}(\tfrac{1}{3})^{17}$$
$$+\frac{20\,!}{4\,!\,16\,!}(\tfrac{1}{3})^{4}(\tfrac{2}{3})^{16}+\frac{20\,!}{5\,!\,15\,!}(\tfrac{1}{3})^{5}(\tfrac{2}{3})^{15}+\frac{20\,!}{6\,!\,14\,!}(\tfrac{1}{3})^{6}(\tfrac{2}{3})^{14}\Big]$$

$$=1-\frac{1}{3^{20}}\Big[2^{20}+(20\times 2^{19})+(190\times 2^{18})+(1140\times 2^{17}(+(4845\times 2^{16})$$
$$+(15504\times 2^{15})+(38760\times 2^{14}\Big]$$

$$=1-\frac{2^{14}}{3^{20}}\Big[64+640+3340+9120+19380+31008+38760\Big]$$

$$=1-\frac{2^{14}}{3^{20}}\Big[102012\Big]=1-0.4793$$

$=0.5207$

PROBLEM 12.18. *What is the probability of having 12 three times in 100 tosses, of the dice ?*

SOLUTION. First we calculate the probability of 12 with two dice :

First die :	6	5	4	3	2	1
Second die :	6	5	4	3	2	1

12 can be obtained in only one way. Total no. of cases $=6\times6=36$.
$\therefore$ Probability of obtaining $12=\frac{1}{36}$

If coming up of 12 is called the event, then according to the problem must take place 3 times. There are hundred tosses and E comes up 3 tim Therefore total number of ways in which 12 can come up 3 times $= {}^{100}C_3$.

Now let us find out the probability of any one of these combinations.

Let the combination be

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 6 \ldots\ldots\ldots\ldots & 99 & 100 \\ E & E & E & F & F \ldots\ldots\ldots\ldots & F & F \end{array}$$

The probability that E takes place $=\frac{1}{36}$

" " " E " .. again $=\frac{1}{36}\cdot\frac{1}{36}=(\frac{1}{36})^2$

" " " E " a third time $\frac{1}{36}\cdot\frac{1}{36}\cdot\frac{1}{36}=(\frac{1}{36})^3$

The probability that E does not take place(*i.e.*—F takes place) $=\frac{35}{36}$

The probability that F takes place two times $=(\frac{35}{36})^2$

The probability that F takes place 97 times $=(\frac{35}{36})^{97}$

$\therefore$ The probability that E takes place 3 times and F takes place 97 tim $=(\frac{1}{36})^3(\frac{35}{36})^{97}$

Since this can happen in ${}^{100}C3$ ways, the required probability $= {}^{100}C$ $(\frac{1}{36})^3(\frac{35}{36})^{97}$
$=0.2257.$

PROBLEM 12.19. *What is the probability of having 12 points with tw dice at least three times in hundred throws ?*

SOLUTION. We have seen that the probability of obtaining 12 poi
=

Now, probability of obtaining 12 points 3 times
$= {}^{100}C_3(\frac{1}{36})^3(\frac{35}{36})^{97}$

Similarly " " " 4 times $= {}^{100}C_4(\frac{1}{36})^4(\frac{35}{36})^{96}$

so on ………… …………………

" " " 99 " $= {}^{100}C_{99}(\frac{1}{36})^{99}(\frac{35}{36})^1$

" " " 100 " $= {}^{100}C_{100}(\frac{1}{36})^{100}(\frac{35}{36})^0$

$\therefore$ Probability of having 12 points at least three times

$$= \sum_{x=3}^{100} {}^{100}C_1 \left(\frac{1}{36}\right)^x \left(\frac{35}{36}\right)^{100-x}$$

$$= 1 - \sum_{x=0}^{2} {}^{100}C_1 \left(\frac{1}{36}\right)^x \left(\frac{3}{36}\right)^{100-x}$$

$$= -2\left[\left(\frac{35}{36}\right)^{100} + 100\left(\frac{1}{36}\right)^{100}\left(\frac{35}{36}\right)^{99} + 4950\left(\frac{1}{36}\right)^2\left(\frac{35}{36}\right)^{98}\right]$$

$==0.52$ (Approx).

PROBLEM 12.20. *A bag contains 5 tickets numbered 1, 2, 5. The ticke are drawn one by one without replacement and if the ticket numbered x appears i the xth drawing, this is denoted as a match. Find the probability of having a least one match.*

SOLUTION. Let Ax be the probability that there will be a match in the xth drawing.

Then the required probability of at least one match $=P(A_1+A_2+\ldots A_5)$

Now we know that

$$P(A_1+A_2+\ldots+A_5)=P(A_1)+P(A_2)+\ldots+P(A_5)$$
$$-P(A_1A_2)-P(A_1A_3)\ldots\ldots P(A_4A_5)$$
$$+P(A_1A_2A_3)+P(A_1A_2A_4)+\ldots+P(A_3A_4A_5)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$-P(A_1A_2\ldots A_5)$$

To evaluate this probability, let us calculate these probabilities :

5 tickets can be arranged in 5 ! ways which is the number of all possible cases.

(1) Evaluation of $P(A_1)+P(A_2)+\ldots P(A_5)$

If suppose ticket numbered 1 appears at the first draw, then the other 4 tickets can appear in 4 ! ways. Therefore, the number of favourable cases is 4 !

$$\therefore P(A_1)=\frac{4\,!}{5\,!}$$

Now this can happen with $A_2, A_3 \ldots A_{10}$ $\therefore$ The probability $= {}^5C_1\frac{4\,!}{5\,!}$

(2) Evaluation of $-P(A_1A_2)-P(A_1A_3)\ldots\ldots P(A_4A_5)$

Suppose tickets numbered 1 and 2 appear at the first and second draws respectively. Then the other 3 tickets can appear in 3 ! ways.

Further 2 tickets can be selected out of five tickets in ${}^5C_2$ ways

$$\therefore -P(A_1A_2)-P(A_1A_3)\ldots-P(A_4A_5)=-{}^5C_2\frac{3\,!}{5\,!}$$

Similarly

$$P(A_1A_2A_3)+\ldots\ldots+P(A_3A_4A_5)={}^5C_3\times\frac{2\,!}{5\,!}$$

$$P(A_1A_2A_3A_4)+\ldots\ldots+P(A_2A_3A_4A_5)={}^5C_4\times\frac{1\,!}{5\,!}$$

$$P(A_1A_2A_3A_4A_5)={}^5C_5\times\frac{0\,!}{5\,!}$$

$\therefore$ The required probability $=P(A_1+A_2+A_3+A_4+A_5)$

$$={}^5C_1\frac{4\,!}{5\,!}-{}^5C_2\frac{3\,!}{5\,!}+{}^5C_3\frac{2\,!}{5\,!}-{}^5C_4\frac{1\,!}{5\,!}+{}^5C_2\frac{0\,!}{5\,!}$$

$$=\frac{5}{1}\times\frac{4\,!}{5\,!}-\frac{5.4}{1.2}\times\frac{3\,!}{5\,!}+\frac{5.4.3}{3.2.1}\times\frac{2\,!}{5\,!}-\frac{5.4.3.2}{1.2.3.4}\times\frac{1\,!}{5\,!}$$

$$+\frac{0}{5\,!}=1-\frac{1}{2\,!}+\frac{1}{3\,!}-\frac{1}{4\,!}+\frac{1}{5\,!}$$

$$=1-0.5000+0.1667-0.0417+0.0083$$

$$=1.1750-0.5417=0.6333$$

PROBABILITY OF HYPOTHESIS

PROBLEM 12.21. *The contents of urn 1, 2, 3 are as follows :*

urn 1	*1 white*	*2 black*	*3 red balls*
urn 2	*2 white*	*1 black*	*1 red balls*
urn 3	*4 white*	*5 black*	*3 red balls*

One urn is chosen at random and two balls drawn. They happen to be white and red, what is the probability they come from urn 2 or 3 ?

SOLUTION. The event A is that two balls taken from the selected urn were of white and red colour respectively. This fact can be accounted by three hypotheses.

(*i*) The selected urn was 1, (*ii*) It was 2, and (*iii*) It was 3.

Since the urns are exactly similar, the probabilities of these hypothes before anything is known about the event A are equal. If we call these hypotheses B_1, B_2, and B_3 and denote their probabilities by (B_1), (B_2) (B_3) respectivel then

$$(B_1)=(B_2)=(B_3)-\tfrac{1}{3}$$

The probabilities of A, assuming these hypotheses are.

$$(A,B_1)=\frac{(AB_1)}{(B_1)}=\frac{\frac{1}{5}\times\frac{1}{3}}{\frac{1}{3}}=-\frac{1}{5}$$

$$(A,B_2)=\frac{(AB_2)}{(B_2)}=\frac{\frac{1}{3}\times\frac{1}{3}}{\frac{1}{3}}=\frac{1}{3}$$

$$(A,B_3)=\frac{(AB_3)}{(B_3)}=\frac{\frac{2}{11}\times\frac{1}{3}}{\frac{1}{3}}=\frac{2}{11}$$

$$-(B_1,A)=\frac{(B_1)\,(A,B_1)}{(B_1)(A,B_1)+(B_2)(A,B_2)+(B_3)(A,B_3)}$$

$$=\frac{\frac{1}{5}\times\frac{1}{3}}{(\frac{1}{5}\times\frac{1}{3})+(\frac{1}{3}\times\frac{1}{3})+(\frac{1}{3}\times\frac{2}{11})}=\frac{\frac{1}{3}}{\frac{1}{5}+\frac{1}{3}+\frac{2}{11}}=\frac{165}{118}\times\frac{1}{5}$$

$$=\frac{33}{118}$$

Similarly,
$$(B_2,A)=\frac{\frac{1}{3}\times\frac{1}{3}}{(\frac{1}{5}\times\frac{1}{3})+(\frac{1}{3}\times\frac{1}{3})+(\frac{1}{3}\times\frac{2}{11})}\quad\frac{\frac{1}{3}}{\frac{1}{5}+\frac{1}{3}+\frac{2}{11}}$$

$$=\frac{165}{118}\times\frac{1}{3}=\frac{55}{118}$$

and
$$(B_3,A)=\frac{\frac{1}{5}\times\frac{1}{3}}{(\frac{1}{5}+\frac{1}{3})+(\frac{1}{3}\times\frac{1}{3})+(\frac{1}{3}\times\frac{2}{11})}=\frac{\frac{2}{11}}{\frac{1}{5}+\frac{1}{3}+\frac{2}{11}}$$

$$=\frac{165}{118}\times\frac{2}{11}=\frac{30}{118}$$

Therefore the probability that the balls were taken from urn 2 or urn 3 is $\frac{6}{18}$ or $\frac{30}{118}$

PROBLEM 12.22. *An urn containing 6 balls was filled in the following manner : A coin was tossed 6 times,andlaccordinglyas it showed heads or tails one white or one black ball was put into the urn. Balls are drawn from this urn one at a time, 6 times in succession with replacement. What is the probability that urn contains only white balls ?*

SOLUTION. In this problem event A is that in 6 independent trials only white balls appear. The probability of this is definite but unknown. This fact can be accounted by six hypotheses regarding the number of white balls in the urn. They are that the number of white balls in the urn is either 1, or 2, or 3, or 4, or 5, or 6.

The *a'* priori probability of 1, or 2, or 3, or 4, or 5, or 6 white balls is known by the manner in which the urn is filled. And we find that it is the same as the probability of having 1, or 2, or 3, or 4, or 5, or 6 heads in 6 throws.

These probabilities are respectively $\frac{6!}{1!\,5!}$, $\frac{1}{2^6}$, $\frac{6!}{2!\,4!}$, $\frac{1}{2^6}$, $\frac{6!}{3!\,3!}$, $\frac{1}{2^6}$, $\frac{6!}{4!\,2!}$, $\frac{1}{2^6}$, $\frac{6!}{5!\,1!}$, $\frac{1}{2^6}$ and $\frac{6!}{6!}$, $\frac{1}{2^6}$ or $\frac{6}{64}$, $\frac{15}{64}$, $\frac{20}{64}$, $\frac{15}{64}$ and $\frac{1}{64}$

Assuming the hypothesis 1, the probability of A $=\dfrac{(\frac{1}{6})^6\times(\frac{6}{64})}{\frac{6}{64}}=\left(\dfrac{1}{6}\right)^6$

" " " " 2, " " " A $\dfrac{(\frac{2}{6})^6\times\frac{15}{64}}{\frac{15}{64}}=\left(\dfrac{2}{6}\right)^6$

" " " " 3, " " " A $\dfrac{(\frac{3}{6})^6\times\frac{20}{64}}{\frac{20}{64}}=\left(\dfrac{3}{6}\right)^6$

" " " " 4, " " " A $\dfrac{(\frac{4}{6})^6\times\frac{15}{64}}{\frac{15}{64}}=\left(\dfrac{4}{6}\right)^6$

" " " " 5, " " " A $\dfrac{(\frac{5}{6})^6\times\frac{6}{64}}{\frac{6}{64}}=\left(\dfrac{5}{6}\right)^6$

" " " " 6, " " " A $\dfrac{(\frac{6}{6})^6\times\frac{1}{64}}{\frac{1}{64}}=\left(\dfrac{6}{6}\right)^6=1$

We have to find out the probability that all 6 balls are white, *i.e.* probability of the 6th hypothesis.

This probability $=\dfrac{1\times\frac{1}{64}}{\left(\frac{6}{64}\times(\frac{1}{6})^6\right)+\left(\frac{15}{64}\times(\frac{2}{6})^6\right)+\left(\frac{20}{64}\times(\frac{3}{6})^6\right)+\left(\frac{15}{64}\times(\frac{4}{6})^6\right)+\left(\frac{6}{64}\times(\frac{5}{6})^6\right)+\left(\frac{1}{64}\times(\frac{6}{6})^6\right)}$

$$=\frac{1}{\frac{1}{6^6}\left(6+960+14580+61440+156250+46656\right)}$$

$$=\frac{46656}{279892}$$

PROBLEM 12.23. *From an urn containing 3 white and 5 black balls, 4 balls are transferred into an empty urn.. From this urn 2 balls are taken and they both happen to be white. What is the probability hat the third ball from the same urn will be white ?*

SOLUTION. When we draw two balls from the urn, they are white. Therefore the urn must contain at least 2 white balls. Now regarding the colour of the 4 balls transferred from the first urn to the second urn, we have the following hypothesis

B1 : 2 white balls and 2 black balls.

B2 : 3 white balls and 1 black ball

A priori probability of the first hypothesis$=(B_1)=\dfrac{{}^3C_2\times{}^5C_2}{{}^8C_4}$

$$=\frac{3\times5\times4\times4\times3\times2}{2\times8\times7\times6\times5}=\frac{3}{7}$$

A priori " " second " $=(B_2)=\dfrac{{}^3C_3\times{}^5C_1}{{}^8C_4}$

$$=\frac{1\times5\times2\times3\times4}{8\times7\times6\times5}=\frac{1}{14}$$

The event A is the fact that both the balls are white.

Assuming the first hypothesis the probability of $A=(A,B_1)=\dfrac{(AB_1)}{B_1}$

$$=\frac{\frac{1}{6}\times\frac{3}{7}}{\frac{3}{7}}=\frac{1}{6}$$

" " second " " " " $A=(A,B_2)=\dfrac{(A.B_2)}{B_2}$

$$=\frac{\frac{3}{6}\times\frac{1}{14}}{\frac{1}{14}}=\frac{3}{6}$$

About the future event—drawing a ball from the urn—we can distinguis two possibilities (*i*) The balls are returned to the urn, (*ii*) They are not returned.

(*i*) Since the balls are returned to the urn, C becomes independent of A as soon as it is known which one of the hypothesis has materialised.

$$\text{Therefore } (C,AB_1)=(C,B_1)=\frac{(C,B_1)}{(B_1)}=\frac{\frac{2}{4}\times\frac{3}{7}}{\frac{3}{7}}=\tfrac{1}{2}$$

$$\text{and } (C,AB_2)=(C,B_2)=\frac{CB_2}{B_2}=\frac{\frac{3}{4}\times\frac{1}{14}}{\frac{1}{14}}=\frac{3}{4}$$

$$\text{Now, } (C,A)=\frac{(B_1)\,(A,B_1)\,(C,AB_1)+(B_2)\,(A,B_2)\,(C,AB_2)}{(B_1)\,(A,B_1)+(B_2)\,(A,B_2)}$$

$$=\frac{\left(\frac{3}{7}\times\frac{1}{6}\times\frac{1}{2}\right)+\left(\frac{1}{14}\times\frac{3}{6}\times\frac{3}{4}\right)}{\left(\frac{3}{7}\times\frac{1}{6}\right)+\left(\frac{1}{14}\times\frac{3}{6}\right)}$$

$$=\frac{\frac{1}{28}+\frac{3}{112}}{\frac{1}{14}+\frac{1}{28}}=\frac{\frac{7}{112}}{\frac{3}{28}}=\frac{7}{112}\times\frac{28}{3}=\frac{7}{12}$$

The probability that third ball will be white$=\frac{7}{12}$

(*ii*) If the two balls are not returned, we have

$(C,AB_2)=0$ —for there were only two white balls.

$$\text{and } (C,AB_2)=\frac{(C)(AB_2)}{(AB_2)}=\frac{\frac{1}{2}\left(\frac{3}{6}\times\frac{1}{14}\right)}{\left(\frac{3}{6}\times\frac{1}{14}\right)}=\tfrac{1}{2}$$

$$\therefore\qquad (CA)=\frac{\left(\frac{3}{7}\times\frac{1}{6}\times 0\right)+\left(\frac{1}{14}\times\frac{3}{6}\times\frac{1}{2}\right)}{\left(\frac{3}{7}\times\frac{1}{6}\right)+\left(\frac{1}{14}\times\frac{3}{6}\right)}$$

$$=\frac{\frac{1}{56}}{\frac{9}{84}}=\frac{84}{9\times 56}=\frac{1}{6}$$

MATHEMATICAL EXPECTATION

PROBLEM 12.24. *A bag contains 3 blue and 4 white counters. A blue counter is worth 5 paisa coin and a white counter is worth 1 rupee. A man i allowed to draw three counters at random. What is the value of his expectation?*

SOLUTION. Expectation $E(x)=p_1 x_1+p_2 x_2+\ldots\ldots pnxn$, where p_1, p_2 $\ldots pn$ are probabilities of $x_1, x_2 \ldots xn$.

(i) Possible sums	(ii) value Rs.	 P	(iii) Probability	(iv) px
3 coins + 0 rupee	—	15	$\frac{{}^3C_3 \times {}^4C_0}{{}^7C_3} = \frac{1}{35} = \frac{1}{35}$	$\frac{15 \times 1}{35}$
2 coins + 1 rupee	1	10	$\frac{{}^3C_2 \times {}^4C_1}{{}^7C_3} = \frac{3 \times 4}{35} = \frac{12}{35}$	$\frac{110 \times 12}{35}$
1 coin + 2 rupees	2	5	$\frac{{}^3C_1 \times {}^4C_2}{{}^7C_3} = \frac{3 \times 6}{35} = \frac{18}{35}$	$\frac{205 \times 18}{35}$
0 coin + 3 rupees	3	0	$\frac{{}^3C_0 \times {}^4C_3}{{}^7C_3} = \frac{1 \times 4}{35} = \frac{4}{35}$	$\frac{300 \times 4}{35}$

value of the expectation $= \left(\frac{15}{35} + \frac{1320}{35} + \frac{3690}{35} + \frac{1200}{35}\right)$ paisa

$= \frac{6225}{35}$ Paisa$= \frac{1245}{7}$ or Paisa 177.6=Rs. 1.776

PROBLEM 12.25. *There are ten counters in a bag, 6 of them worth 5/- each. The other 4 are of equal but unknown value. If the expectation from drawing a single counter at random is 4/-, find the unknown value.*

SOLUTION. Let the equal but unknown value of counters be x.

Then, the probability of drawing from the 6 counters of value 5/-$=\frac{6}{10}$ and probability of drawing from 4 counters is $\frac{4}{10}$

Value of the expectation—(5/-)$\frac{6}{10}$+(x/-)$\frac{4}{10}$

The value of the expectation is known to be 4/-

$$\frac{5 \times 6}{10} + \frac{4x}{10} = 4$$

$$\text{or } 4x = 10$$

$$x = 2.5$$

PROBLEM 12.26. *The expectation of the number of failures preceding the first success in an indefinite set of independent trials, with constant probability p of success, is*

$qp + 2q^2p + 3q^3p + \cdots\cdots \quad + \frac{pq}{(1-q)_2} = \frac{q}{p}$

SOLUTION.

Trials	1	2	3	4	5	6........
Probability of success in this trial (p)	p	qp	q^2p	q^3p	q^4p	q^5p
No. of failures preceding the success in this trial (x)	0	1	2	3	4	5........
Px	0	qp	$2q^2p$	$3q^3p$	$4q^4p$	$5q^5p$

$\therefore$Expectation $=0+1qp+2q^2p+3q^3p+5q^4p+5q^5p+\cdots\cdots$

$=qp(1+2q+3q^2+4q^3+\cdots\cdots)$

Let S $=1+2q+3q^2+4q^3+\cdots\cdots$

$\therefore qs \quad = \quad q+2q^2+3q^3+\cdots\cdots$

S$-qs \quad =1+q+q^2+q^3+\cdots\cdots$

$$= \frac{1}{1-q}$$

$$S\,(1-q) = \frac{1}{1-q} \text{ or } S = \frac{1}{(1-q)^2}$$

$$\therefore \text{Expectation} = \frac{pq}{(1-q)^2} \qquad = \frac{q}{p}$$

PROBLEM 12.27. *What is the mathematical expectation of the sum of points on 10 dice.*

SOLUTION. Let the number of point on first die be x_1, on second die x_2, on third die $x_3 \ldots .$, on 9th die x_9 and on 10th die x_{10}.

Then the sum of the points$=S=x_1+x_2+\ldots+x_{10}$

Now, $E\,(s)=E(x_1)+E(x_2)+\ldots\ldots\ldots E(x_{10})$

Now for one die the expectation

$$=(\tfrac{1}{6}\times 1)+(\tfrac{1}{6}\times 2)+(\tfrac{1}{6}\times 3)+(\tfrac{1}{6}\times 4)+(\tfrac{1}{6}\times 5)+(\tfrac{1}{6}\times 6)$$

$$= \frac{21}{6} = \frac{7}{2}$$

$$\therefore (Es) = \frac{10\times 7}{2} = 35$$

PROBLEM 12.28. ***Balls are taken one by one out of an urn containing 10 white balls and 8 black balls until the first white ball is drawn. What is the expectation of the number of black balls preceding the first white ball ?***

SOLUTION.

No. of trials	1	2	3	8	9
Probability of taking out a white ball	$\frac{10}{10}$	$\left(\frac{8}{18}\times\frac{10}{7}\right)$	$\left(\frac{8}{18}\times\frac{7}{17}\times\frac{10}{16}\right)\left(\frac{8}{18}\times\frac{7}{17}\times\frac{6}{16}\times\frac{10}{15}\right)\cdots$	$\left(\frac{8\times 7\times..2}{18\times 17..12}\times\frac{1}{12}\right)$	$\left(\frac{8\times 7\times..1}{18\times 17+..11}\times\frac{10}{10}\right)$
No. of black balls preceding the white balls	0	1	2	3.........................7	8

$$\therefore \text{Expectation} = \left(0\times\frac{10}{10}\right)+\left(1\times\frac{8\times 10}{18\times 17}\right)+\left(2\times\frac{8\times 7\times 10}{18\times 17\times 16}\right)+\ldots+\left(7\times\frac{8\times 7..\ 2\times 10}{18\times 17\times 12\times 11}\right)+1$$

$$\left(8\times\frac{8\times 7\times ..\ 1\times 10}{18\times 17\times ..11\times 10}\right)$$

$$=\frac{10}{10}\left[\frac{8}{17}+2\times\frac{8\times 7}{17\times 16}+3\times\frac{8\times 7\times 6}{17\times 16\times 15}+..+7\times\frac{8\times 7..\ 2}{17\times 16..11}+8\times\frac{8\times 7..\ 1}{17\times 16..10}\right]$$

$$=\frac{10}{18}\left[\frac{8}{17}+\frac{7}{17}+\frac{21}{17\times 15}+\frac{2}{17}+\frac{10}{13\times 17}+\frac{3}{13\times 17}+\frac{7}{11\times 13\times 17}+\frac{4}{11\times 12\times 17\times 5}\right]$$

$$=\frac{10}{18}\times\frac{1224}{11\times 17\times 5}=\frac{8}{11}$$

REVISIONAL PROBLEMS—12

PROBLEM 12.29 (*a*) The probability that a certain event happened was 1/10, and A, who is accurate in 49 cases out of 50, said that it happened. B agrees with A stating that the event happened; B is accurate in a cases out of 10. What is the probability that it actually did occur ?

(*b*) Suppose if C, who is accurate in 7 cases out of 10, denies that the event mentioned above happened, what is the probability that it happend ?

(*Ans* . (*a*) 49/50 (*b*) 21/22)

PROBLEM 12.30. Five coins whose faces are marked 2, 3 are thrown. What is the chance of obtaining a total of 12 ?

(*Ans*: 5/16)

PROBLEM 12.31. A and B throw one dice for a prize of Rs. 165 which is to be won by the player who first throws six. If A has the first throw, what are their respective expectations ?

(*Ans*. A=Rs. 90 & B=Rs. 75)

PROBLEM 12.32. (a) During war one ship in 10 was sunk on the average in making a certain voyage. What was the probability that at least 3 out of a convoy of 6 ships would arrive safely ?

(b) Three houses of the same type were advertised to be let in a locality. Three men made separate applications for a house. What is the probability :

(i) that all the three made applications for the same house;

(ii) that each of the three applied for a different house; and

(iii) that two of them applied for the same house and the third for one of the other houses ?

(*Ans*. (a) 9. 18, 730/10,00,000
(b) (i) 1/9
(ii) 2/9
(iii) 2/3)

PROBLEM 12.33. Three newspapers A, B and C are published in a certain city. It is estimated from a survey of the adult population:

20% read A, 16% read B, 14% read C, 8% read both A and B, 5% read both A and C, 4% read both B and C, 2% read all three. What percentage reads at least one of the papers ? Of these that read at least one, what percentage reads both A and B ?

(*Ans*. 35%; 28%)

PROBLEM 12.34. A, B and C in order lose a coin. The first one to throw a head wins. What are their respective chances to win ?

(*Ans*. A=4/7: B=2/7: C=1/7)

PROBLEM 12.35. There are 10 tickets, 4 of which are blanks, the others being marked 1, 2, 3, 4, 5 and 6 respectively. Find the probability of obtaining total of 6 in three trials, one ticket being drawn and replaced on each occasion.

(*Ans*. .063)

PROBLEM 12.36. From a bag containing 200 black balls and 100 white balls, a ball is drawn at random. What is the probability that it will be white ?

(*Ans* : 1/3)

PROBLEM 12.37. Two dice are rolled simultaneously. What is the probability that

(i) sixes appear on both of them;

(ii) sixes appear on none of them ?

(*Ans*. 1/36 and 35/36)

PROBLEM 12.38. There are two ways to enter a house and three ways to go out. In how many ways can you enter and leave the house ?

(*Ans* : 6)

PROBLEM 12.39. From a group of seven men and four ladies, a committee is to be formed. If there are to be six people on the committee, in how many ways can the committee be composed :

(i) if there are to be exactly two ladies on the committee.

(ii) if there are to be at least two ladies on the committee.

(*Ans* : (i) 210; (ii) 371)

PROBLEM 12.40. Eight students appear in a class-test. There are three prizes for the first three places. In how many ways can the prizes be distributed ?

(*Ans* : 336)

Theoretical Frequency Distributions

13

Distributions which are not obtained by actual observation or experiment bu are mathematically deducted on certain assumptions are called theoretical frequency distributions.

Three such distributions are of special importance. Their equations are given below—

Binomial Distribution. The equation of this distribution is $(p+q)^n$ and in chapters of probability its expansion has already been discussed and illustrated.

Poisson Distribution. The equation of this distribution is

$$P(x)=\frac{m^x}{x\,!}\,e^{-m} \text{ where } m=np.$$

Normal Distribution. The equation of this distribution is

$$P(x)=\frac{1}{\sigma\sqrt{2\pi}}\;e-\frac{(x-\bar{x})^2}{2\sigma^2}\text{, where}$$

e is the base of natural logarithms and has value of 2.71828 and $\sqrt{2\pi}=2.5066$.

BINOMIAL DISTRIBUTION

PROBLEM 13.1. *Twelve coins are tossed. What are the probabilities in a single tossing of getting (i) 9 or more heads, (ii) less than 3 heads, (iii) at least 8 heads? If the 12 coins are tossed 4096 times, on how many occasions would you expect these to be (iv) less than 3 heads, (v) at least 2 heads, (vi) exactly 6 heads?*

SOLUTION. The probability of getting head$=\frac{1}{2}$ and that of tail$=1-\frac{1}{2}=\frac{1}{2}$. The B. D. is given by $(\frac{1}{2}+\frac{1}{2})^{12}=(\frac{1}{2})^{12}+{}^{12}C_{11}\,(\frac{1}{2})^{11}\,(\frac{1}{2})+\ldots+(\frac{1}{2})^{12}$

(*i*) Probability of getting 9 or more heads :

,, ,, ,, ,, 12 heads$=(\frac{1}{2})^{12}$

,, ,, ,, ,, 11 heads$={}^{12}C_{11}\,(\frac{1}{2})^{11}\,(\frac{1}{2})$

,, ,, ,, ,, 10 heads$={}^{12}C_{10}\,(\frac{1}{2})^{10}\,(\frac{1}{2})^{2}$

,, ,, ,, ,, 9 heads$={}^{12}C_{9}\,(\frac{1}{2})^{9}\,(\frac{1}{2})^{3}$

$\therefore$ Probability of getting 9 or more heads$=(1)^{12}(1+12+66+220)$

$$=\frac{299}{4096}=0.073.$$

(*ii*) Probability of getting 0 head$=(\frac{1}{2})^{12}$

,, ,, ,, ,, 1 ,, $={}^{12}C_1(\frac{1}{2})^{11}(\frac{1}{2})$

,, ,, ,, ,, 2 ,, $={}^{12}C_2(\frac{1}{2})^{10}(\frac{1}{2})^{2}$

$\therefore$ Probability of getting less than 3 heads$=(1)^{12}\{1+12+66\}$

$$=\frac{79}{4096}=0.0193.$$

(*iii*) Probability of getting 9 or more heads$=0.073$

" " " " exactly 8 " $={}^{12}C_8(\frac{1}{2})^8(\frac{1}{2})^4=0{\cdot}1208$

" " " " at least 8 " $=0.073+0.1208$

$=0.1938.$

(*iv*) If 12 coins are tossed 4096 times, less than 3 heads will appear

$4096 \times \frac{79}{4096}$ times$=79$

(*v*) The probability of getting 0 and 1 head is$=(\frac{1}{2})^{12}+12\times(\frac{1}{2})^{12}=\frac{13}{4096}$

$\therefore$ Probability of getting at least two heads$=1-\frac{13}{4096}=\frac{4083}{4096}$

$\therefore$ The number of occasions when at least two heads will appear in 4096 tosses of 12 coins

$$=4096 \times \frac{4083}{4096}=4083$$

(*vi*) The probability of getting exactly six heads

$$={}^{12}C_6\times(\tfrac{1}{2})^6+(\tfrac{1}{2})^6=\frac{924}{4096}$$

$\therefore$ In 4096 tosses exactly six heads will come up in

$$4096 \times \frac{924}{4096}=924 \text{ times.}$$

PROBLEM 13.2. *Four dice are thrown and the number of sixes in each throw are recorded. This is repeated 108 times. Write down the theoretical frequencies of 0, 1, 2, 3, and 4 sixes. Calculate the mean number of sixes in each throw.*

SOLUTION. The probability of a six in the throw of a dice$=\frac{1}{6}$

If 4 dice are thrown, the probability distribution is given by :

$$(\tfrac{5}{6}+\tfrac{1}{6})^4=(\tfrac{5}{6})^4+{}^4C_1\,(\tfrac{5}{6})^3\,\tfrac{1}{6}+{}^4C_2\,(\tfrac{5}{6})^2\,(\tfrac{1}{6})^2+{}^4C_1(\tfrac{5}{6})\,(\tfrac{1}{6})^0+(\tfrac{1}{6})^4$$

$$=\left(\frac{625}{1296}\right)+\left(4\times\frac{125}{1296}\right)+\left(6\times\frac{25}{1296}\right)+\left(4\times\frac{5}{1296}\right)+\left(\frac{1}{1296}\right)$$

Theoretical frequency of 0, 1, 2, 3, 4 sixes in 108 trials is given by :

Number of sixes (x)	$P(x)$	Theoretical frequency $108\times P(x)=f$
0	$\frac{625}{1296}$	$52\frac{1}{12}$
1	$\frac{125}{324}$	$41\frac{2}{3}$
2	$\frac{25}{216}$	$12\frac{1}{2}$
3	$\frac{5}{324}$	$1\frac{2}{3}$
4	$\frac{1}{1296}$	$\frac{1}{12}$
		$\Sigma f=108$

Number of sixes in each throw :—

$$\begin{array}{lcl} & & fx \\ 0\times \frac{625}{12} & = & 0 \\ 1\times \frac{125}{3} & = & \frac{125}{3} \\ 2\times \frac{25}{2} & = & 25 \\ 3\times \frac{5}{3} & = & 5 \\ 4\times \frac{1}{12} & = & \frac{1}{3} \\ \hline & & \Sigma fx = \frac{216}{3} \end{array}$$

$\therefore$ Mean number of sixes in each throw $=\frac{\Sigma fx}{\Sigma f}=\frac{216}{3\times 108}=\frac{2}{3}$

PROBLEM 13.3. *In a packet of flower seeds $\frac{2}{5}$ are known to be pink flowerin and the remainder are yellow. Calculate the probability of getting 0, 1, 2,......6 pi flowers in a row of six plants. If 250 rows each of 6 plants are planted, approximate how many will contain (a) all pink flowers, (b) all yellow flowers ?*

SOLUTION. The probability of getting pink flowers$=\frac{2}{5}$; and the probabili of getting yellow flowers$=1-(\frac{2}{5})=\frac{3}{5}$.

The probability of getting 0, 1,.........6 pink flowers is given by the binomi distribution $(\frac{2}{5}+\frac{3}{5})^6$

$=(\frac{3}{5})^6+{}^6C_1\,(\frac{3}{5})^5\,(\frac{2}{5})+{}^6C_2\,(\frac{3}{5})^4\,(\frac{2}{5})^2+{}^6C_3\,(\frac{3}{5})^3(\frac{2}{5})^3+{}^6C_4\,(\frac{3}{5})^2\,(\frac{2}{5})^4+{}^6C_5(\frac{3}{5})\,(\frac{2}{5})^5+(\frac{2}{5})^6$

$=\frac{729}{15625}+6\,\frac{486}{15625}+15\,\frac{324}{15625}+20\,\frac{216}{15625}+15\,\frac{144}{15625}+6\,\frac{96}{15625}+\frac{64}{15625}$

$\therefore$ The probabilities of getting 0, 1,.........6 pink flowers are

$\frac{729}{15625}$, $\frac{2916}{15625}$, $\frac{4860}{15625}$, $\frac{4320}{15625}$, $\frac{2160}{15625}$, $\frac{576}{15625}$, and $\frac{64}{15625}$

In 250 rows the number of rows containing all pink flowers

$=\frac{64}{15625}\times 250=\frac{640}{625}=1$ (approx).

In 250 rows the number of rows containing all yellow flowers

$=\frac{729\times 250}{15625}=\frac{7290}{625}=12$ (approx)

PROBLEM 13.4. *A card was drawn from a pack of playing cards and the replaced and the pack shuffled. This was repeated ten times and the number of blac suit cards was noted. A thousand results were obtained in this way and are given here The number of black cards is denoted by x.*

x :	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Total*
f :	2	8	*46*	*116*	*211*	*243*	*208*	*119*	*40*	*7*	*0*	*1000*

What distribution would be expected to apply ? Calculate the theoretical frequencies and compare with the experimental frequencies.

SOLUTION. The number of black suit cards in a pack of 52 cards is 26. Therefore the probability of getting a black card is $\frac{1}{2}$

Since cards are drawn 10 times, the probability distribution is given by $P(x) = (\frac{1}{2}\times\frac{1}{2})^{10}$

$$=(\tfrac{1}{2})^{10}+{}^{10}C_1(\tfrac{1}{2})^9\tfrac{1}{2}+{}^{10}C_2(\tfrac{1}{2})^8(\tfrac{1}{2})^2+\ldots\ldots{}^{10}C_2(\tfrac{1}{2})(\tfrac{1}{2})^0+(\tfrac{1}{2})^{10}$$

The theoretical frequencies :

No. of black card	$P(x)$	Theoretical frequencies approx. $1000+P(x)$
0	$\frac{1}{210}=\frac{1}{1024}$	1
1	$\frac{10}{210}=\frac{10}{1024}$	10
2	$\frac{45}{210}=\frac{45}{1024}$	44
3	$\frac{120}{210}=\frac{120}{1024}$	117
4	$\frac{210}{210}=\frac{210}{1024}$	205
5	$\frac{252}{210}=\frac{252}{1024}$	246
6	$\frac{210}{210}=\frac{210}{1024}$	205
7	$\frac{120}{210}=\frac{120}{1024}$	117
8	$\frac{45}{210}=\frac{45}{1024}$	44
9	$\frac{10}{210}=\frac{10}{1024}$	10
10	$\frac{1}{210}=\frac{1}{1024}$	1

PROBLEM 13.5. *The following data, due to W. F. R. Weldon shows the results of throwing 12 dice 4096 times, throw of 4, 5 or 6 being called a success—*

Success	*Frequency*	*Successes*	*Frequency*
0	—	*7*	*847*
1	*7*	*8*	*536*
2	*60*	*9*	*257*
3	*198*	*10*	*71*
4	*430*	*11*	*11*
5	*731*	*12*	—
6	*948*		
		Total	*4096*

Find the expected frequencies, and compare the actual mean and standard de
tion with those of the expected distribution.

SOLUTION. The probability of a success$=\frac{3}{8}=\frac{1}{2}$

The expected frequencies will be given by the B.D.

$$4096\left(\tfrac{1}{2}+\tfrac{1}{2}\right)^{12}=4096\left[\left(\tfrac{1}{2}\right)^{12}+{}^{12}C_1\left(\tfrac{1}{2}\right)^{11}\left(\tfrac{1}{2}\right)\times{}^{12}C_2\left(\tfrac{1}{2}\right)^{10}\left(\tfrac{1}{2}\right)^2+\ldots\ldots+\left(\tfrac{1}{2}\right)^{12}\right]$$

Success	Expected frequency	Success	Expected frequency
0	$4096=(2)^{12}$	7	$792\times1=792$
1	$12\times1=\ 12$	8	$495\times1=495$
2	$66\times1=\ 66$	9	$220\times1=220$
3	$220\times1=220$	10	$66\times1=\ 66$
4	$495\times1=495$	11	$12\times1=\ 12$
5	$792\times1=792$	12	$1=\ 1$
6	$924\times1=924$	Total	4096

Calculation of the actual mean and standard deviation.

x	f	d	df	d^2f
0	—	−6	0	0
1	7	−5	− 35	175
2	60	−4	− 240	960
3	198	−3	− 594	1782
4	430	−2	− 860	1720
5	731	−1	− 731	731
6	948	0	0	0
7	847	1	847	847
8	536	2	1072	2144
9	257	3	771	2313
10	71	4	284	1136
11	11	5	55	275
12	—	6	0	0
			+ 569	12083

$$\text{Mean}\ =6+\frac{569}{4096}=6.13\dot{9}$$

$$\sigma=\frac{1}{\mathrm{N}}-\left(\Sigma\, fd-d\right)^2$$

$$=\frac{12083}{4096}-(0.139)^2$$

$$=2.9499-0.0193$$

$$\therefore\ a=\sqrt{\ 2.9306}=1.71$$

Mean of the expected distribution $=np=12\times\frac{1}{8}=6$ Standard deviation of the expected distribution$=\sqrt{npq}=\sqrt{12\times\frac{1}{2}\times1\times\frac{1}{2}}=\sqrt{3}=1.73$

PROBLEM 13.6 *Assuming that 50% of the population of a town takes tea and assuming that 100 investigators each take 10 individuals to find out if they take tea, how many investigators would you expect to report that three people or less take tea ?*

SOLUTION. The probability of an individual taking tea or $p=\frac{1}{2}$
As such $q=\frac{1}{2}$

We will have to raise a binomial $100\left(\frac{1}{2}+\frac{1}{2}\right)^{10}$

The number of investigators who are expected to report that 0, 1, 2, and 3 people take tea would be equal to—

$$N\ (q^{10})+(N10q^9p)+(N45q^8p^2)+(N120q^7p^3)$$

$$=\frac{100}{1{,}024}+\frac{1{,}000}{1{,}024}+\frac{4{,}500}{1{,}024}+\frac{12{,}000}{1{,}024}=17$$

PROBLEM 13.7 *An irregular six-faced die is thrown, and the expectation that n 10 throws it will give five even numbers is twice the expectation that it will give ur even numbers. How many times in 10,000 sets of 10 throws would you expect it o give no even numbers?*

SOLUTION. Let the probability of getting an even number be p and that of not getting $1-p=q$

∴The probability distribution is given by $(p+q)^{10}$

∴The probability of getting 5 even numbers $={}^{10}C_5 p^5 q^5$

and „ „ „ 4 „ „ $={}^{10}C_4 p^4 q^6$

Now, ${}^{10}C_5 p^5 q^5 = 2 \times {}^{10}C_4 p^4 q^6$

$$\text{or } \frac{10\times9\times......6}{5\times4\times......1} p^5 q^5 = 2 \times \frac{10\times9\times8\times7}{4\times3\times......1} p^4 q^6$$

or $\frac{6}{8} p = 2q$ or $6p = 10(1-p)$ or $16q = 10$

or $p = \frac{5}{8}$ and $q = 1 - \frac{5}{8} = \frac{3}{8}$

If there are 1,000 sets of 10 throws, the binomial distribution will be $10{,}000 = 10{,}000\ (\frac{5}{8}+\frac{3}{8})^{10}$

The probability of getting zero even numbers is

$${}^{10}C_3\ (\tfrac{5}{8})^0(\tfrac{3}{8})^{10} = 1 \times 1 \times (0.375)^0 = 0.00005495$$

∴ Expected number of times when no even numbers are obtained

$$= 10{,}000 \times 0.00005495 = 0.5495 = 1 \text{ (Approx)}.$$

PROBLEM 13.8 *Find the avercge and variance in terms of p, q, n for the binomial distribution.*

$$P(r) = \frac{n!}{r!\ (n-r)\ !} p^r q^{n-r}$$

Compute these to the nearest whole number from the following data, and hence obtain appropriate values of p, q, n.

r	*P(r)*	*r*	*P(r)*	*r*	*P(r)*
2	*0.001*	*9*	*0.136*	*16*	*0.016*
3	*0.004*	*10*	*0.140*	*17*	*0.008*
4	*0.013*	*11*	*0.127*	*18*	*0.004*
5	*0.030*	*12*	*0.103*	*19*	*0.002*
6	*0.055*	*13*	*0.075*	*20*	*0.001*
7	*0.087*	*14*	*0.050*		
8	*0.117*	*15*	*0.030*		

SOLUTION. Calculation of mean and standard deviation :

r	*P(r)*	$r \times P(r)$	$r^2P(r)$
2	0.001	0.002	0.004
3	0.004	0.012	0.036
4	0.013	0.052	0.208
5	0.030	0.150	0.750
6	0.055	0.330	1.980
7	0.087	0.609	4.263

$\Sigma P(r) = 0.999$

$\Sigma rP(r) = 9.996$

$\Sigma r^2P(r) = 108.008$

$$\text{Mean} = \frac{9.996}{0{\cdot}999}$$

8	0.117	0.936	7.488
9	0.136	1.224	11.016
10	0.140	1.400	14.000
11	0.127	1.397	15.367
12	0.103	1.236	14.832
13	0.075	0.976	12.688
14	0.050	0.700	9.800
15	0.030	0.450	6.750
16	0.016	0.256	4.096
17	0.008	0.136	2.312
18	0.004	0.072	1.296
19	0.002	0.038	0.722
20	0.001	0.020	0.400

Mean $=10.006=10$

$a^2 = s^2 - d^2$

$=108.11-100=8.11$

$=8$

Now we know :

$np=10$; $npq=8$ or $np(1-p)=1$ or $np-np^2=8$ or $np^2=2$

or $10p=2$ or $p=\frac{1}{5}$ $\therefore q=1-p=\frac{4}{5}$

$\therefore n\frac{1}{5}=10$ or $n=10$; $p=\frac{1}{5}$, $q=\frac{4}{5}$ and $n=10$

PROBLEM 13.9. *Derive expressions for the first four moments of the binomial distribution.*

SOLUTION. The binomial distribution is given by the terms of expansion of $(q+p)^n$, where $q+p=1$.

No. of successes (x)	0	1	2	3.........n
Probability $P(x)$	q^n	${}^nC_1 q^{n-1}p$	${}^nC_2 q^{n-2}p^2$	${}^nC_3 q^{n-3}p^3$......p^n

(*i*) First moment about zero, or the mean :

$$\delta_1=\frac{1}{\Sigma P(x)}\,\Sigma xP(x) \text{ or } \Sigma P(x)=(p+q)^n-1$$

$$\Sigma x\, P(x)=0\times q^n+1\times nC_1 q^{n-1}p+2\times {}^nC_2 q^{n-2}p^2+3\times {}^nC_3 q^{n-3}p^3+\ldots+nxp^n$$

$$=0+1\times nq^{n-1}p+2\times\frac{n(n-1)}{2\,!}q^{n-2}p^2+3\times\frac{n(n-1)(n-2)}{3\,!}q^{n-3}p^3+\ldots+n\times p^n$$

$$=np\,[q^{n-1}+(n-1)\,q^{n-2}p^2\frac{(n-1)(n-2)}{2\,!}q^{n-3}p^3+\ldots+p^{n-1}$$

$$=np\,(q+p)^{n-1}=np$$

$\therefore \delta_1=$ Mean $=np$.

(*ii*) Second moment about zero.

$$\delta_2=\frac{1}{\Sigma P(x)}\,\Sigma x^2\,P(x)=\Sigma x^2\,P(x)$$

$$=\Sigma\,\{x(x-1)\times{}^nC_x q^{n-x}p^x\} \text{ for } x^2=x^2-x+x$$

$$=\Sigma x(x-1)\times{}^nC_x q^{n-x}p^x+\Sigma x\times{}^nC_x q^{n-x}p^x$$

$$=\Sigma x(x-1)\times\frac{n(n-1)\ldots\ldots(n-x+1)}{x(x-1)(x-2)\ldots2\times1}q^{n-x}p^x+\Sigma x\times\frac{n(n-1)\ldots(n-x+1)}{x(x-1)\ldots2-1}q^{n-x}p^x$$

$$=n(n-1)\,p^2\Sigma\,\frac{(n-2)(n-3)\ldots(n-x+1)}{(x-2)(x-3)\ldots\ldots2\times1}\,q^{n-x}p^{x-2}+np\Sigma\,\frac{(x-1)\ldots(n-x+1)}{(x-1)\ldots2\times1}\,q^{n-x}p^{x-1}$$

$$=n(n-1)p^2\,[\Sigma\,{}^{n-2}C_{x-2}q^{n-x}p^{x-2}]+np\Sigma\,{}^{n-1}C_{x-1}q^{n-x}p^{x-1}$$
$$n(n-1)\,p^2\,(q+p)^{n-2}+np\,(q+p)^{n-1}=n(n-1)\,p^2+np$$

(*iii*) Third moment about zero.

$$\delta_3=\frac{1}{\Sigma P(x)}\,\Sigma x^2P(x)=\Sigma x^3\,P(x)$$

$$\delta^3=\Sigma\,[x(x-1)(x-2)+3x(x-1)+x]\,{}^nC_xq^{n-x}p^x,\text{ for } x^3=x(x-1)(x-2)+3x(x-1)+x$$

$$=\Sigma x(x-1)(x-2)\times\frac{n(n-1)\ldots\ldots(n-x\times1)}{x(x-1)(x-2)\ldots2\times1}q^{n-x}p^x+\Sigma\,3x(x-1)$$

$$\frac{n(n-1)\ldots(n-x+1)}{x(x-1)\ldots2\times1}\,q^{n-x}p^x+\Sigma x\,\frac{n(n-1)\ldots(n-x+1)}{x(x-1)\ldots\ldots2\times1}\,q^{n-x}p^x$$

$$=n(n-1)(n-2)\,p^3\Sigma\,\frac{(n-3)\ldots(n-x+1)}{(x-3)\ldots\ldots2\times1}\,q^{n-x}p^{x-3}+3n(n-1)p$$

$$\Sigma\,\frac{(n-2)\ldots(n-x+1)}{(x-2)\ldots2\times1}\,q^{n-x}p^{x-2}+np\Sigma\,\frac{(n-1)\ldots\ldots(n-x+1)}{(x-1)\ldots\ldots2\times1}\,q^{n-x}p^{x-1}$$
$$=n(n-1)(n-2)p^3\,(q+p)^{n-3}+3n(n-1)\,p^2\,(q+p)^{n-2}+np(q+p)^{n-1}$$
$$=n(n-1)(n-2)p^3+3n(n-1)p^2+np$$

(*iv*) Fourth moment about zero :

$$\delta_4=\frac{1}{\Sigma P(x)}\,\Sigma x^nP(x)=\Sigma x^nP(x)$$

$$\delta_4=\Sigma\,[x(x-1)(x-2)(x-3)+x(x-1)(x-2)+7x(x-1)+x]\,{}^nC_xq^{n-x}p^x$$
$$\text{for}\quad x^4=x(x-1)(x-2)(x-3)+6x(x-1)(x-2)+7x(x-1)+x$$

$$=n(n-1)(n-2)(n-3)pn\Sigma\,\frac{(n-4)\ldots\ldots(n-x+1)}{(x-4)\ldots\ldots\ldots2\times1}\,q^{n-x}p^{x-4}$$

$$+6n(n-1)(n-2)p^3\,\Sigma\,\frac{(n-3)\ldots\ldots(n-x+1)}{(x-3)\ldots\ldots\ldots3\times2\times1}\,q^{n-x}p^{x-3}$$

$$+7n(n-1)p^2\Sigma\,\frac{(n-2)\ldots\ldots(n-x+1)}{(x-2)\ldots\ldots\ldots2\times1}\,q^{n-x}p^{x-2}$$

$$+np\,\frac{(n-1)\ldots\ldots(n-x+1)}{(x-1)\ldots\ldots\ldots2\times1}\,q^{n-x}p^{x-1}$$
$$=n(n-1)(n-2)(n-3)\,p^4+6n(n-1)(n-2)\,p^3+7n(n-1)\,p^2+np$$

(*v*) Moments about Mean :

$$\mu_2=\delta_2-\delta_1{}^2=n(n-1)\,p^2+np-n^2p^2$$
$$=np-np^2=np\,(1-p)$$
$$=npq$$

$$\mu_3=\delta_3-3\delta_1\delta_2+2\delta_1{}^3$$
$$=n(n-1)(n-2)p^3+3n(n-1)p^2+np-3np\,[n(n-1)p^2+np]+2n^3p^3$$
$$=(n^3-3n^2+2n)\,p^3+(3n^2-3n)\,p^2+np-3n^3p^3+3n^2p^2-3n^2p^2+2n^3p$$

$$=2np^3-3np^2+np=2np^3-2np^2-np^2+np$$
$$=2np^2(p-1)-np(p-1)$$
$$=npq-2np^2q$$
$$=npq(1-2p)=npq(q-p)$$
$$\mu_4=\delta_4-4\delta_1\delta_3+6\delta_1^2\delta_2-3\delta_1^4$$
$$=[n(n-1)(n-2)(n-3)p^4+6n(n-1)(n-2)p^3+7n(n-1)p^2+np]$$
$$-4np[n(n-1)(n-2)]p^3+3n(n-1)p^2+np)+6n^2p^2[n(n-1)p^2+np]-3n^4p^4$$
$$=[(n^4-6n^3+11n^2+6n)p^4+)6n^3-18n^2+12n)p^3+(7n^2-7n)p^2+np]$$
$$=[(4n^4-12n^3+8n^2)p^4+(12n^3-12n^2)p^3+4n^2p^2]+[(6n^4-6n^3)p^4+6n^3p^3]-3n^4p^4$$
$$=(3n^2p^4+6np^4)+(-6n^2p^3+12np^3)+(3n^2p^2-7np^2)+np$$
$$=np^4(3n-6)+np^3(-6n+12)+np^2(3n-7)+np$$
$$=3np^4(n-2)-np^3(n-2)+np^2(3n-6)+np-np^2$$
$$=3np^4(n-2)-6np^3(n-2)+3np^2(n-2)+np(1-p)$$
$$=3np^2(n-2)\{p^2-2p+1\}+np(1-p)$$
$$=3n(n-2)p^2(p-1)^2+np(1-p)=3n(n-2)p^2+q^2+npq$$
$$=3n^2p^2q^2-6np^2q^2+npq=3n^2p^2q^2+npq(1-6pq)$$

PROBLEM 13.10. *Show that if two symmetrical binomial distributions of degree n (and of the same number of observations) are so superseded that the rth term of the one coincides with the $(r+1)$ the term of the other, the distribution formed by adding superseded terms is a symmetrical binomial of degree $(n+1)$.*

SOLUTION. Let the first distribution be $N=N(\frac{1}{2}+\frac{1}{2})^n$ and the second distribution with the same number of observations be $N=N(\frac{1}{2}+\frac{1}{2})^n$

(For the symmetrical binomial distribution $p=q=\frac{1}{2}$).

The terms of the first and second distribution are given by :

First distribution—$N=N\left(\frac{1}{2^n}+{}^nC_1\frac{1}{2^n}+{}^nC_2\frac{1}{2^n}+r\ldots{}^nC_r\frac{1}{2^n}+\cdots\frac{1}{2^n}\right)$

Second distribution—$N=N\left(\frac{1}{2^n}+{}^nC_1\frac{1}{2^n}+{}^nC_2\frac{1}{2^n}+{}^nC_3\frac{1}{2^n}\right.$
$$\left.+\ldots+{}^nC_{r+1}\frac{1}{2^n}+\cdots\frac{1}{2^n}\right)$$

The distribution formed by adding superseded terms is : $2N=N$

$$\frac{1}{2^n}[1+(1+{}^nC_1)+({}^nC_1+{}^nC_2)+\ldots\ldots+({}^nC_r+{}^nC_{r-1})+\ldots\ldots+1]$$

Now $${}^nC_r+{}^nC_{r+1}=\frac{n(n-1)(n-2)\ldots\ldots(n-r+1)}{r(r-1)(r-2)\ldots\ldots2\times1}+\frac{n(n-1)\ldots\ldots(n-r)}{(r+1)r(r-1)\ldots2\times1}$$
$$=\frac{n(n-1)(n-2)\ldots\ldots(n-r+1)}{r(r-1)(r-2)\ldots\ldots2\times1}\left(1+\frac{n+r}{r+1}\right)$$
$$=\frac{(n+1)}{(r+1)}\times\frac{n(n-1)\ldots\ldots(n-r+1)}{r(r-1)\ldots\ldots2+1}={}^{n+1}C_{r+1}$$
$$\therefore {}_2N=\frac{N}{2n}[1+{}^{n+1}C_1+{}^{n+1}C_2+\ldots\ldots+{}^{n+1}C_{r+1}+\ldots\ldots+1]$$
or $$N=\frac{N}{2n+1}[1+{}^{n+1}C_1+{}^{n+1}C_2+\ldots\ldots+{}^{n+1}C_{r+1}+\ldots\ldots+1]$$
or $N=N(\frac{1}{2}+\frac{1}{2})^{n+1}$

This is a symmetrical binomial distribution of degree $(n+1)$.

POISSON DISTRIBUTION

PROBLEM 13.11. *A skilled typist, on routine work, kept a record of mistakes made per day during 300 working days.*

Mistakes per day :	*0*	*1*	*2*	*3*	*4*	*5*	*6*
No. of days	*143*	*90*	*42*	*12*	*9*	*3*	*1*

Compare the frequencies of the Poisson series which has the same total frequency and mean as the above distribution.

SOLUTION. Calculation of the mean :

Mistakes per day (x)	Number of days (f)	(xf)
0	143	0
1	90	90
2	42	84
3	12	36
4	9	36
5	3	15
6	1	6
	300	$\Sigma xf=267$

$\therefore$ Mean$=m=0.89$

The Poisson distribution is $P(x)=\frac{mx}{x\,!}\,e^{-m}=\frac{(0.89)}{x\,!}\,e^{-.89}$.

Mistakes per day		$P(x)$	Theoretical Frequency $300\times P(x)$	Frequency (nearest whole number)
0	$\frac{(0.89)^0}{0\,!}\,e^{-.89}=1\times e^{-.89}$	$=0.4107$	123.21	123
1	$\frac{(0.89)^1}{1\,!}\,e^{-.89}=0.89\times 0.4107$	$=0.3655$	109.65	110
2	$\frac{(0.89)^1}{2\,!}\,e^{-.89}=\frac{0.89}{2}\times 0.3655$	$=0.1626$	48.78	49
3	$\frac{(0.89)^3}{3\,!}\,e^{-.89}=\frac{0.89}{3}\times 0.1626$	$=0.0482$	14.46	14
4	$\frac{(0.89)^4}{4\,!}\,e^{-.89}=\frac{0.89}{4}\times 0.0482$	$=0.0107$	3.21	3
5	$\frac{(0.89)^5}{5\,!}\,e^{-.86}=\frac{0.89}{5}\times 0.0107$	$=0.0019$	0.57	1
6	$\frac{(0.89)^6}{6\,!}\,e^{-.89}=\frac{0.89}{6}\times 0.0019$	$=0.0003$	0.09	—
				300

PROBLEM 13.12. *Assuming that the average number of defective articles in a certain manufacturing process is 3% find the theoretical frequency of samples of fifty with 0, 1, 2, 3......defectives.*

SOLUTION. $p=\frac{3}{100},\; n=50 \quad \therefore m=1.5$

The Poisson distribution is given by $P(x)=\frac{m^x}{x\,!}\,e^{-m}=\frac{1.5}{x}\,e^{-1.5}$

The probability of a sample with *0* defective $=1\times e^{-1.5}=0.2231$

,, ,, ,, ,, 1 ,, $=1.5\,e^{-1.5}=0.3347$

,, ,, ,, ,, 2 ,, $=\frac{1.5^2}{2\,!}\,e^{-1.5}=0.2510$

,, ,, ,, ,, 3 ,, $=\frac{1.5^3}{3\,!}\,e^{-1.5}=0.1255$

,, ,, ,, ,, 4 ,, $=\frac{1.5^4}{4\,!}\,e^{-1.5}=0.0471$

,, ,, ,, ,, 5 ,, $=\frac{1.5^5}{5\,!}\,e^{-1.5}=0.0141$

,, ,, ,, ,, 6 ,, $=\frac{1.5^6}{6\,!}\,e^{-1.5}=0.0035$

,, ,, ,, ,, 7 ,, $=\frac{1.5^7}{7\,!}\,e^{-1.5}=0.0007$

If there were hundred samples of fifty each, the theoretical frequencies of 0, 1, 2...... defectives would be 22.3, 33.5, 25.1, 12.6, 4.7, 1.4 and 0.4.

PROBLEM 13.13. *Assuming that one in 80 births is a case of twins, calculate the probability of two or more sets of twins in a certain town on a day when 30 births occur. Compare the results obtained by using (i) the binomial series and (ii) Poisson's approximation.*

SOLUTION. (*i*) By using the binomial series :

The probability of a set of twins $=\frac{1}{80}$

Hence the binomial expansion is given by $\left(\frac{1}{80}+\frac{79}{80}\right)^{30}$

Probability of getting 0 set of twins $=\left(\frac{79}{80}\right)^{30}=(0.9875)^{30}$

Probability of getting 1 set of twins $={}^{30}C_1\left(\frac{79}{80}\right)^{29}\left(\frac{1}{80}\right)$

$$=30\times(0.9875)^{29}\times(0.0125)$$

Calculation :

$\log(0.9875)^{30}=30\ \log(0.9875)=30\times\bar{1}.9945=\bar{1}.8350$

og$\{30+(0.9875)^{29}+(0.0125)\}=\log\ 30+29\ \log\ (0.9875)+\log\ 0.0125$

$$=1.4771+\bar{1}.8405+\bar{2}.0969=\bar{1}.4145$$

Probability of getting 0 set of twins Anti log $\bar{1}.8350=0.6839$

,, ,, ,, ,, 1 ,, ,, $=$Anti log $\bar{1}.4145=0.2597$

,, ,, ,, ,, less than 2 ,, ,, $=0.6839+0.2597=0.9436$

,, ,, ,, two or more sets of twins $=1-0.9436=0.0564$

(*ii*) By using Poisson's approximation :

Poisson's approximation is $\frac{m^x}{x\,!}e\text{-}m$, where $m=np$.

Now $n=30$ and $p=\frac{1}{80}$ $m=np=\frac{3}{8}=0.375$

Probability of getting 0 set of twins$=\frac{(0.375)^0}{0\,!}e^{-.375}=e^{-.375}=0.6873$

,, ,, ,, 1 ,, ,, $=\frac{(0.375)^1}{1\,!}e^{-.375}=0.375+0.6873$

$=0.2577$

,, ,, less than 2 sets of twins $=0.6875+0.2577=0.9450$

,, ,, at least 2 sets of twins$=1-0.9450=.055$

PROBLEM 13.14. *If the average number of rejects in the manufacture of a certain article is 4%, what are the probabilities of 0, 1, 2, 3, 4 rejects in a sample of 10 articles taken at random ?*

SOLUTION. $p=\frac{4}{100}$, $n=10$ $\therefore$ $m=np=0.4$

No. of Rejects	$P(x)=\frac{m^x}{x\,!}e\text{-}m$	
0	$\frac{(0.4)^0}{0\,!}e^{-.4}=1\times0.6703$	$=0.6703$
1	$\frac{(0.4)}{1\,!}e^{-.4}=0.4\times0.6703$	$=0.2681$
2	$\frac{(0.4)^2}{2\,!}e^{-.4}=0.08\times0.6703$	$=0.0536$
3	$\frac{(0.4)^3}{3\,!}e^{-.4}=0.1065\times0.6703$	$=0.0071$
4	$\frac{(0.4)^4}{4\,!}e^{-.4}=0.01065\times0.6703$	$=0.0007$

The probabilities of 0, 1, 2, 3, and 4 rejects are 0.6703, 0.2681, 0.0536, 0.0071, and 0.0007 respectively.

PROBLEM 13.15. *Fit a Poisson's distribution to the set of observations :—*

Deaths	*0*	*1*	*2*	*3*	*4*
Frequency	*122*	*60*	*15*	*2*	*1*

and calculate the theoretical frequencies.

SOLUTION.

Mean of the distribution—

$$m=\frac{(122\times0)+(60\times1)+(15\times2)+(2\times3)+(1\times4)}{122+60+15+2+1}=\frac{100}{200}=.5$$

Poisson's equation is

$$P\,(x)=\frac{m^x}{x\,!}e\text{-}m=\frac{.5^x}{x\,!}e^{-.5}$$

Theoretical frequencies— Theoretical Frequency

0 Death or $P(_0)=\frac{(.5)^0}{0\,!}e^{-.5}=1 \quad \times e^{-.5} \quad = \quad {\cdot}61 \quad =.61 \quad \times 200=122$

1 Death or $P(_1)=\frac{(.5)^1}{1\,!}e^{-.5}=\frac{.5}{1}\times .61 \quad = \quad .305 \quad =.305 \quad \times 200=61$

2 Death or $P(_2)=\frac{(.5)^2}{2}e^{-.5}=\frac{.5}{2}\times .305 \quad = \quad .07625=.07625\times 200=15$

Similarly for 3 deaths the frequency will be = 2

and for 4 deaths the frequency will be = 0

PROBLEM 13.16. *Certain mass produced articles, of which 0.9 per cent are defective, and packed in cartons each containing 100. What proportion of cartons are free from defective articles, and what proportion contain 2 or more defectives ?*

SOLUTION. $p=.005$, $n=100 \quad \therefore \; m=0.5$

The Poisson distribution is $P(x)=\frac{.5^x}{x\,!}e^{-.5}$

The probability of a carton free from defective articles $=P(o)$

$$=\frac{.5^0}{0\,!}e^{-.5} \qquad =0.6065$$

The probability of a carton with one defective $=P\ (1)$

$$=\frac{.5}{1\,!}e^{-.5} \qquad =0.303$$

The probability of a carton with 2 or more defectives

$$=1—0.6065=0.30325$$
$$=1—0.90975=0.09025$$

Therefore the proportion of cartons free from defective articles is 60.65% and with 2 or more defectives is 9% (approx).

PROBLEM 13.17. *In a certain factory turning razor blades there is a small chance $\frac{1}{500}$ for any blade to be defective. The blades are supplied in packets of 10. Use Poisson's distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.*

SOLUTION.

$$p=\frac{1}{500} \text{ and } n=10 \; \therefore \; m=\frac{10}{500}=.02$$

Poisson's distribution is given by

$$P(^x)=\frac{m^x}{x\,!}e^{-m}=\frac{(.02)^x}{x\,!}e^{-.02}$$

Probability of 0 defectives is

$$P(_0)=\frac{(.02)}{0\,!}e^{-.2}=.9802$$

Hence the theoretical frequency for 0 defective $=.9802\times 10{,}000=9802$

Similarly the theoretical frequency for 1 defective

$$=\left\{\frac{(.02)^1}{1\,!}e^{-.02}\right\}\times 10{,}000 \qquad =196 \text{ and}$$

for 2 defectives

$$=\left\{\frac{(.02)^2}{2\,!}e^{-.02}\right\}\times 10{,}000 \qquad =2$$

PROBLEM 13.18. *In 1,000 extensive sets of trials for an event of small probability the frequencies fi of the numbers xi of success proved to be*

x :	0	1	2	3	4	5	6	7
f :	305	365	210	80	28	9	2	1

Show that the mean number of successes is 1.2, and find out the frequencies of the Poissonian distribution with the same mean and the same total frequency.

SOLUTION. Calculation of the mean :

x	f	fx
0	305	0
1	365	365
2	210	420
3	80	240
4	28	112
5	9	45
6	2	12
7	1	7
	$\Sigma f=1000$	$\Sigma fx=1201$

$\therefore$ Mean $=m=1.2$

The Poissonian distribution is given by $P(x)=\frac{1.2x}{x^1}$ e-1.2

Hence the theoretical frequencies are :

x		
0	$\frac{1.2^0}{0\,!}e^{-1.2}=1\times 0.3012=0.3012$	301.2
1	$\frac{1.2}{1\,!}e^{-1.2}=1.20\times 3012=0.3614$	361.2
2	$\frac{1{\cdot}2^2}{2\,!}e^{-1.2}=0.6\times 0.3614=0.2168$	216.8
3	$\frac{1.2^3}{3\,!}e^{-1.2}=0.4\times 0.2168=0.0867$	86.7
4	$\frac{1.2^4}{4\,!}e^{-1.2}=0.3\times 0.0867=0.0260$	26.0
5	$\frac{1.2^5}{5\,!}e^{-1.2}=0.24\times 0.0260=0.0062$	6.2

6	$\frac{1.2^6}{6!}e^{-1.2}=0.2\times0.0062=0.0012$	1.2
7	$\frac{1.2^7}{7!}e^{-1.2}=0.17\times0.0012=0.0002$	0.2

PROBLEM 13.19. *The following numbers of deaths from cancer of the adren gland were recorded each month in a period of 24 months*

3 4 2 1 3 4 3 2 1 6 3 5 4 2
2 0 4 3 2 6 0 4 0 2

Fit a Poisson curve to the distribution :

SOLUTION. The frequency distribution of the data is given by :

No. of deaths :	0	1	2	3	4	5	6	7 and over	Total
No. of months :	3	2	6	5	5	1	2	0	24

No. of deaths x	No. of months f	fx
0	3	0
1	2	2
2	6	12
3	5	15
4	5	20
5	1	5
6	2	12
	$\Sigma f=24$	$\Sigma fx=66$

$$\therefore m=\frac{66}{24}=2.75$$

The Poisson distribution with the mean 2.75 is given by

$$P(x)=\frac{(2.75)^x}{x!}e^{-2.75}$$

No. of deaths x	$P(x)$	Theoretical frequency $24\times P(x)$
0	$1 \times e^{-2.75} = .064$	1.536
1	$2.75 \times .064 = .176$	4.244
2	$\frac{2.75}{2}\times .176 = .242$	5.808
3	$\frac{2.75}{3}\times .242 = .222$	5.318
4	$\frac{2.75}{4}\times .222 = .1526$	3.662
5	$\frac{2.75}{5}\times .1526 = .0839$	2.014
6	$\frac{2.75}{6}\times .0839 = .0385$	0.924
7 and over	——	0.514

PROBLEM 13.20. *From records of the Prussian army corps kept over twenty years, the following data were obtained showing the number of deaths caused by the kick of a horse. Determine the mean and calculate the theoretical Poisson frequencies.*

Number of deaths per army corps per annum	*Frequency of occurrence*
0	*109*
1	*65*
2	*22*
3	*3*
4	*1*
Total	*200*

SOLUTION. The mean $=\dfrac{(0\times109)+(1\times65)+(2\times22)+(3\times3)+(4\times1)}{200}$

$$=\frac{122}{200}=0.61$$

The Poisson distribution is given by $P(x)=\dfrac{m^x}{x\,!}e^{-m}$

In this example $P(x)=\dfrac{(0.61)^x}{x\,!}e^{-.61}$

$$P(_0)=1\times e\text{-}0.61=1-0.61+\frac{(.61)^2}{2\,!}-\frac{(.61)^3}{3\,!}+\frac{(.61)^4}{4\,!}-\frac{(.61)^5}{5\,!}$$

$$=1-0.61000+0.18605-0.03783+0.00577-0.0070+0.0007$$
$$=1.19189-0.64853=.54340.$$

$$P(_1)=0.61\times.5434 \qquad =.3315$$

$$P(_2)=\frac{(0.61)^2}{2\,!}\times.5434=\frac{.3315\times.61}{2} \qquad =.1011$$

$$P(_3)=\frac{(0.61)^3}{3\,!}\times.5434=\frac{.1011\times.61}{3} \qquad =.0206$$

$$P(_4)=\frac{(0.61)^4}{4\,!}\times.5434=\frac{.0206\times.61}{4} \qquad =.0031$$

$\therefore$ Theoretical frequencies are $200\times(.5434,\ .3315,\ .1011,\ .0206,.0031)=$ 108.68, 66.30, 20.22, 4.12, 0.62.

Poissonian frequencies in whole numbers are : 109, 66, 20, 4, 1.

PROBLEM 13.21. *A car hire firm has two cars, which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson Distribution with mean 1.5.*

Calculate the proportion of days on which neither car is used, and the proportion of days on which some demand is refused.

If each car is used an equal amount, on what proportion of days is one of the cars not in use ? What proportion of demands has to be refused ?

SOLUTION. The distribution of the number of demands is given by

$$P(x)=\frac{(1.5)}{x\,!}e^{-1.5}$$

Demands per day x	Proportion of days $P(x)$
0	$1 \times e^{-1.5} = 0.2231$
1	$1.5 \times 0.2231 = 0.3346$
2	$\frac{1.5}{2} \times 0.3346 = 0.2510$
3	$\frac{1.5}{3} \times 0.2510 = 0.1255$
4	$\frac{1.5}{4} \times 0.1255 = 0.0471$
5	$\frac{1.5}{5} \times 0.0471 = 0.0141$
6	$\frac{1.5}{6} \times 0.0141 = 0.0035$
7	$\frac{1.5}{7} \times 0.0035 = 0.0008$

(*i*) Proportion of days on which neither car is used $= 0.2231$

(*ii*) Proportion of days on which some demand is refused

$$=1-(0.2231+0.3346+0.2510)$$
$$=1-(0.8087=0.1913$$

(*iii*) Proportion of days on which only one car is demanded$=0.3346$

Since each car is used an equal amount, proportion of days on which only one of the cars is not in use $= \frac{0.3346}{2} = 0.1673$

Demands per day	*No. of days per 1000 days*	*Total Demands*	*Demands refused*
1	223	—	—
1	335	335	—
2	251	502	—
3	126	378	126
4	47	188	94
5	14	70	42
6	3	18	12
7	1	7	5
	1000	1498	279

Proportion of demands refused$= \frac{279}{1498} = 0.1862.$

PROBLEM 13.22. *Obtain the values of the first four moments about* (*a*) *zero,*) *mean for the Poisson Distribution.*

SOLUTION. If the value be xi and frequency fi, the moments are given by :

	about zero	about mean
1st Moment	$\delta_1=\frac{1}{n}\Sigma fixi=\bar{x}$	$\mu_1=\frac{1}{n}\Sigma fi\,(xi-\bar{x})=0$
2nd Moment	$\delta_2=\frac{1}{n}\Sigma fixi^2$	$\mu_2=\frac{1}{n}\Sigma fi\,(xi-\bar{x})^2$
3rd Moment	$\delta_3=\frac{1}{n}\Sigma fixi^3$	$\mu_3=\frac{1}{n}\Sigma fi\,(xi-\bar{x})^3$
4th Moment	$\delta_1=\frac{1}{n}\Sigma fixi^4$	$\mu_4=\frac{1}{n}\Sigma fi\,(xi-\bar{x})^4$

The Poisson Distribution is $P(x)=\frac{m^{\bar{x}}}{x\,!}e^{-m}$

(a) Calculation of the first moment about origin or of the means :

No. of successes	0	1	2	3	4………
$P(x)$	$1.e^{-m}$	$\frac{m}{1\,!}e^{-m}$	$\frac{m^2}{2\,!}e^{-m}$	$\frac{m^3}{3\,!}e^{-m}$	$\frac{m^4}{4\,!}e^{-m}$………

$$\Sigma P(x)=1.e^{-m}+\frac{m}{1\,!}e^{-m}+\frac{m^2}{2\,!}e^{-m}+\ldots\ldots\ldots$$

$$=e^{-m}\left(1+\frac{m}{1\,!}+\frac{m^2}{2\,!}+\frac{m^3}{3\,!}+\ldots\ldots\ldots\right)$$

$$=e^{-m}\times e^{-m}=1$$

$$\Sigma P(x)=0+\frac{m}{1\,!}e^{-m}+2\times\frac{m^2}{2\,!}e^{-m}+3\frac{m^3}{3\,!}e^{-m}+\ldots\ldots\ldots$$

$$=Me^{-m}\left(1+m+\frac{m^2}{2\,!}+\frac{m^3}{3\,!}+\ldots\right)=m\times e^{-m}\times e^{m}$$

$$=M.$$

Mean of the Poisson Distribution$=\frac{\Sigma xP(x)}{\Sigma P(x)}=\frac{m}{1}=m.$

(b) Calculation of the second moment about zero.

$$\delta_2=\Sigma x^2P(x)=0+1^2\times\frac{m}{1!}e^{-m}+2^2\frac{m^2}{2\,!}e^{-m}+3^2\frac{m^3}{3\,!}e^{-n}+4^2\frac{m^4}{4\,!}e^{-m}+\ldots$$

$$=\Sigma e^{-m}\times x^2\frac{m^x}{x\,!}$$

Now $x^2=x(x-1)+x$

$$\therefore\ \delta_2=e^{-m}\Sigma\frac{m^x}{x\,!}\left[x(x-1)+x\right]=e^{-m}\Sigma\frac{x(x-1)}{x\,!}m^x+e^{-m}\Sigma\frac{x}{x!}m^x$$

$$=e^{-m}\left[\Sigma\frac{1}{(x-1)\,!}m^x+\Sigma\frac{1}{(x-1)\,!}m^x\right]=e^{m}\left[m^3\Sigma\frac{m^{x-2}}{(x-2)\,!}+m\Sigma\frac{m^{x-3}}{(x-1)\,!}\right]$$

$$=e-[m^2\times e^{-m}+Me^{-m}]=m^2+m$$

(c) Third moment about zero :

$$\delta_3=\Sigma x^3 P(x)=0+1^3\times\frac{m}{1!}e^{-m}+2^3\frac{m^2}{2!}e^{-m}+3^3\frac{m^3}{3!}e^{-m}+\ldots\ldots\ldots$$

$$=\Sigma e^{-m}\frac{m^x}{x!}x^3$$

Now $x^3=x(x-1)(x-2)+3x(x-1)+x$

$$\therefore\ \delta_2=e^{-m}\Sigma\ [x(x-1)(x-2)+3x(x-1)+x]\frac{m^x}{x!}$$

$$=e^{-m}\left[\Sigma\frac{x(x-1)(x-2)}{x!}m^x+\Sigma\frac{3x(x-1)}{x!}m^x\Sigma\frac{x}{x!}m^x\right.$$

$$=e^{-m}\left[\Sigma\frac{m^x}{(x-3)!}+\Sigma\frac{3m^x}{(x-2)!}+\Sigma\frac{m^x}{(x-1)!}\right]$$

$$=e^{-m}\left(m^3\Sigma\frac{m^{x-3}}{(x-3)!}+3m^2\Sigma\frac{m^{x-2}}{(x-2)!}+m\Sigma\frac{m^{x-1}}{(x-1)!}\right)$$

$$=e^{-m}[m^3e^{m}+3m^2e^{m}+me^{m}]$$

$$=m^3+3m+m.$$

(d) Fourth moment about zero :

$$\delta_4=\Sigma x^4P(x)=0+1^4\frac{m}{1!}e^{-m}+2^4\frac{m^2}{2!}e^{-m}+3^4\frac{m^3}{3!}e^{-m}+\ldots\ldots\ldots$$

$$=\Sigma e^{-m}\frac{m^x}{x!}x^4$$

Now $x^4=x(x-1)(x-2)(x-3)+6x(x-1)(x-2)+7x(x-1)+x$

$$\delta_4=e^{-m}\Sigma\left(x(x-1)(x-2)(x-3)+6x(x-1)(x-2)+7x(x-1)+x\right)\frac{m}{x!}$$

$$=e^{-m}\left(\Sigma\frac{x(x-1)(x-2)(x-3)}{x!}m^x+\Sigma\frac{6x(x-1)(x-2)}{x!}m^x+\right.$$

$$\left.\Sigma\frac{x(x-1)}{x!}m^x+\Sigma\frac{x}{x!}m^x\right)$$

$$=e^{-m}\left(\Sigma\frac{m^x}{(x-4)!}+\Sigma\frac{6m^x}{(x-3)!}+\Sigma\frac{7m^x}{(x-2)!}+\Sigma\frac{m^x}{(x-1)!}\right)$$

$$=e^{-m}\left(m^4\Sigma\frac{m^{x-4}}{(x-4)!}+6m^2\Sigma\frac{m^{x-3}}{(x-3)!}+7m^2\Sigma\frac{m^{x-2}}{(x-2)!}+m\Sigma\frac{m^{x-1}}{(x-1)!}\right)$$

$$=e^{-m}[m^4e^{m}+6m^3e^{m}+7m^2e^{m}+me^{m}]$$

$$=m^4+6m^3+7m^2+m$$

(e) Moments about mean :

We know that : $\mu_2=\delta_2-\delta_1^2$

$\mu_3=\delta_2-3\delta_1\delta_2+2\delta_1^2$ $\mu_4=\delta_4-4\delta_1\delta_2+6\delta_1^2\delta_2-3\delta_1^4$

∴ Second moment about mean $=m^2+m-m^2=m$

∴ Third ,, ,, ,, $=m^3+3m^2+m-3m(m^2+m)+2m^3=m$

∴ Fourth ,, ,, ,, $=(m^4+6m^3+7m^2+m)+4m(m^3+3m^2+m)$

$+6m^2(m^2+m)-3m^4=3m^2+m.$

PROBLEM 13.23. *Show that the fourth moment about the mean for the Poisson distribution* $e^{-m}\left(1+\frac{m}{1!}+\frac{m^2}{2!}+\dots\right)$ *is* $\mu=m(1+3m)$.

Hence compute the fourth moment of the following Poisson series.

x	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*Total*
Frequency	*54.335*	*33.145*	*10.110*	*2.050*	*0.315*	*0.040*	*0.005*	*100*

SOLUTION.

x	f	fx
0	54.335	0
1	33.145	33.145
2	10.110	20.220
3	2.050	6.150
4	0.315	1.260
5	0.040	0.200
6	0.005	0.030
	$\Sigma f=100$	$\Sigma fx=61.005$

$$\therefore\ x=m=\frac{61.005}{100}=0.61005$$

$$\mu_4=m(1+3m)=0.61005\ (1+1.83015)=0.726$$

PROBLEM 13.24. *Obtain the two beta co-efficients for the binomial and the Poisson distributions.*

SOLUTION. We know that $\beta_1=\frac{\mu_3^2}{\mu_2^3}$ and $\beta_2=\frac{\mu_4}{\mu_2^2}$

(*i*) For the Binomial distribution

$\mu_2=npq$, $\mu_3=npq\ (q-p)$ and $\mu_4=3n^2p^2q^2+npq(1-6pq)$

$$\therefore\ \beta_1=\frac{n^2p^2q^2\ (q-p)^2}{n^3p^3q^3}=\frac{(q-p)^2}{npq}$$

$$\beta_2=\frac{3n^2p^2q^2+npq\ (1-6pq)}{n^2p^2q^2}=3+\frac{1-6pq}{npq}$$

(*ii*) For the Poisson distribution :

$\mu_2=m$, $\mu_3=m$ and $\mu_4=3m^2+m$

$$\beta_1=\frac{m^2}{m^3}=\frac{1}{m}\quad\text{and}\quad\beta_2=\frac{3m^2+m}{m^2}=3+\frac{1}{m}$$

NORMAL DISTRIBUTION

PROBLEM 13.25. *Find out the equation of the normal curve which has the same mean, standard deviation and total frequency as that of the following observed distribution :*

Successes	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*Total*
Frequency	—	*7*	*60*	*198*	*430*	*731*	*948*	*847*	*536*	*257*	*71*	*11*	—	*4096*

SOLUTION. Calculation of the mean and standard deviation :

x	f	d	df	d^2f
0	—	-6	0	0
1	7	-5	-35	175
2	60	-4	-240	960
3	198	-3	-594	1782
4	430	-2	-860	1720
5	731	-1	-731	731
			-2460	
6	948	—	0	0
7	847	1	847	147
8	536	2	1072	2144
9	257	3	771	2313
10	71	4	284	1136
11	11	5	55	275
12	—	6	0	0
			$+3029$ $+569$	12083

$$\text{Mean}=6+\frac{569}{4096}=6.139$$

$$\sigma^2=\frac{12083}{4096}-(0.139)^2=2.9499-0.0193$$

$$=2.9306 \quad \therefore \sigma=1.71$$

The equation of the normal curve is :—

$$y=\frac{N}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\overline{x})^2}$$

on substituting the values we have :

$$y=\frac{4096}{1.71\sqrt{2\pi}}e^{-\frac{1}{5.8612}(x-6.14)^2}$$

$$=954.8\times e^{-\frac{1}{59.712}(x-6.14)^2}$$

PROBLEM 13.26. *Calculate the frequencies of the normal distribution which has the same mean, standard deviation and total frequency as the distribution given below :*

x :	60—	65—	70—	75—	80—	85—	90—	95—	Total
f :	3	21	150	335	326	135	26	4	=1000

SOLUTION. First, we have to calculate the mean and the standard deviation. We find that the mean is $\overline{x}=79.945$ and the S. D. or $\sigma=5.445$

Next we calculate the frequencies :

Since the normal distribution extends from $-\infty$ to $+\infty$, the lower limit 60 is replaced by $-\infty$ and the upper limit is $+\infty$.

Calculation

(1) Interval limit	(2) $\frac{x-\bar{x}}{\sigma}$	(3) Area to right	(4) Proportion of area	(5) Frequency Col.(4)×1000
$-\infty$	$-\infty$	1.0000	0.0031	3.1
65—	2. 74	0.9969	0.0308	30.8
70—	1.826	0.9661	0.1480	148.0
75—	0.908	0.8181	0.3221	322.1
80—	0.011	0.4960	0.3119	319.1
85—	0.928	0.1769	0.1444	144.4
90—	1.846	0.0321	0.0296	29.6
95—	2. 76	0.0029	0.0296	2.9
$+\infty$	$+\infty$	0.0000	—	—
				1000.0

Col. (2)

(i) $\frac{-\infty-\bar{x}}{\sigma}=-\infty,$

(ii) $\frac{65-79.945}{5.445}=2.74$

(iii) $\frac{70-79.945}{5.445}=1.826$

(iv) $\frac{75-79.945}{5.445}=0.908$

(v) $\frac{80-79.945}{5.445}=0.011$

(vi) $\frac{85-79.945}{5.445}=0.928$

(vii) $\frac{90-79.945}{5.445}=1.84$

(viii) $\frac{95-79.945}{5.445}=2.76$

Col. (3)—From Tables

Col. (4)—By subtraction

Area between $-\infty$ and 65=1.0000—0.9969=0031.

Area between 65 and 70=0.9969—0.9661=0.0308

PROBLEM 13.27. *Fit a normal frequency curve to the following distribution :*

:	10—	12—	14—	16—	18—	20—	22—	24—	26—	Total
:	4	30	106	206	272	219	120	37	6	=1000

SOLUTION. The mean of the observed distribution is x=19.14 and standard eviation σ=2.89.

he Calculation :

(1) Interval limit	(2) $\frac{x-\bar{x}}{\sigma}$	(3) Area to right	(4) Proportion of Area	(5) Frequency of the normal curve (4)×1000 ·
$-\infty$	$-\infty$	1.0000	0.0068	6.8
12—	−2.4706	0.9932	0.0308	30.8
14—	−1.7785	0.9624	0.1011	101.1
16—	−1.0865	0.8613	0.2079	207.9
18—	−0.3945	0.6534	0.2704	270.4
20—	+0.2976	0.3830	0.2218	221.8
22—	−0.9896	0.1612	0.1149	114.9
24—	+1.6817	0.0463	0.0375	37.5
26—	+2.3737	0.0088	0.0088	8.8
$+\infty$	$+\infty$	0.0000	—	—
				1000.0

Calculation :

Col. (2) (*i*) $\frac{\infty-19.14}{2.89}=-\infty$

(*ii*) $\frac{12-19.14}{2.89}=2.4706$

(*iii*) $\frac{14-19.14}{2.89}=1.7785$

(*iv*) $\frac{16-19.14}{2.89}=1.0865$

(*v*) $\frac{18-19.14}{2.89}=0.3945$

(*vi*) $\frac{20-19.14}{2.89}=0.2976$

(*vii*) $\frac{22-19.14}{2.89}=0.9896$

(*viii*) $\frac{24-19.14}{2.89}=1.6817$

(*ix*) $\frac{26-19.14}{2.89}=2.3737$

Col. (3) From Tables

Col. (4) Area between $-\infty$ and $12=1.0000-0.9932=.0068$

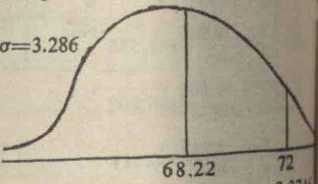
PROBLEM 13.28. *Assume the mean height of soldiers to be 68.22 in., wit variance of 10.8 (in). How many soldiers in a regiment of 1000 would you expect be over six feet tall ?*

SOLUTION. $x=68.22$ and $\sigma^2=10.8, \sigma=3.286$

The standardised deviation

$$\frac{x-\bar{x}}{\sigma}=\frac{72-68.22}{3.286}$$

$=1.15$



Area between $x=0$ and $\bar{x}=1.15$ in the standard normal distribution$=0.374$

$\therefore$ Area to the right of ordinate at $1.15=0{\cdot}5-0.3746=0{\cdot}1254$

$\therefore$ The probability that a soldier is taller than $6'=0.1254$

In a regiment of 1,000 soldiers, the expected frequency$=1000\times0.1254$

$=125.4=125$ soldiers.

PROBLEM 13.29. *Calculate the ordinates of the binomial 1,024* $(0.5+0.5)$ *and compare them with those of the normal curve.*

SOLUTION. The binomial is $1024\left(\frac{1}{2}+\frac{1}{2}\right)^{10}$

$$=1024\ \frac{1}{2^{10}}(1+{}^{10}C_1+{}^{10}C_2+\ldots\ldots+1)=(1+{}^{10}C_1+\ldots\ldots+1)$$

The ordinates of the binomial are $1, {}^{10}C_1, {}^{10}C_2, \ldots {}^{10}C_3 \ldots\ldots 1$ *i.e.* 1, 10, 45, 1 210, 252, 210, 120, 45, 10, 1......................

The mean of this distribution is 5 and the standard deviation is

$$\sqrt{\frac{10}{4}}=\frac{3.162}{2}=1.581$$

The equation of the normal where mean$=5$ and $\sigma=1.581$ is

$$r=\frac{1024}{1.581\sqrt{2\pi}}\ e^{-\frac{(x-5)^2}{5}}$$

From the tables :

(*i*) x	(*ii*) Standard deviation	(*iii*) ordinate for standard normal distribution	(*iv*) ordinate $\frac{1024}{1.581}\times$ Col. (*iii*)
0	$\frac{0-5}{1.581}=3.160$	.0027	1.7
1	$\frac{1-5}{1.581}=2.530$	.0163	10.5
2	$\frac{2-5}{1.581}=1.898$	.0669	42.7
3	$\frac{3-5}{1.581}=1.265$	0.1792	116.1
4	$\frac{4-5}{1.581}=0.632$	0.3265	211.5
5	$\frac{5-5}{1.581}=0.000$	0.3989	258.4
6	$\frac{6-5}{1.581}=0.632$	0.3265	211.5

PROBLEM 13.30. *The mean of a normal frequency distribution of 1000 items is 25 and its standard deviation 2.5 Find the ordinates of the curve at (i) 26.5, (ii) 21, (iii) 19.5, (iv) 32 without using tables. What is the maximum ordinate ?*

SOLUTION. The equation of the normal curve.

$$y=\frac{1000}{2.5\sqrt{2\pi}}e-\frac{(x-25)^2}{12.5}$$

values of x	$e-\frac{(1-2)}{12.5}$	$y=\frac{1000}{.5\sqrt{2\pi}}\times$ Col. (*ii*)
26.5	$e-\frac{2\cdot25}{12.5}=0.833$	133.3
21	$e-\frac{16}{12.5}=e^{-1.28}=0.2770$	44.2
19.5	$e-\frac{30\cdot25}{12.5}=e^{-2.42}=0.0889$	142.0
32	$e-\frac{42}{12.5}=e^{-3.92}=0.198$	3.2

maximum ordinate will be for $x=25$

$$\text{or } y\,\text{max}=\frac{1000}{2.5\sqrt{2\pi}}\,e^0=\frac{1000}{2.5\sqrt{2\pi}}$$

$$=\frac{400}{\sqrt{2\pi}}=400\times0.3989=159.56$$

PROBLEM 13.31. *If two normal populations have the same total frequency, the σ of one is k times that of the other, show that the maximum frequency of the is $\frac{1}{k}$ times of that of the other.*

SOLUTION. Let the total frequency of the first population be N and standard deviation σ.

∴ The total frequency of the second population will be N and its standa deviation $\frac{\sigma}{k}$.

The first normal distribution is given by :

$$y=\frac{N}{\sigma\sqrt{2\pi}}e.\ -\frac{x^2}{2\sigma^2}$$

and the second normal distribution is given by :

$$y=\frac{kN}{\sigma\sqrt{2\pi}}e-\frac{k^2x^2}{\sigma^2}$$

In a normal distribution the maximum frequency corresponds to that value the variate which is equal to mean value, *i.e.*, $x=m$

∴ For the first distribution :

$$y\ \max=y_1=\frac{N}{\sqrt{2\pi}}e^0=\frac{N}{\sigma\sqrt{2\pi}}$$

∴ For the second distribution :

$$y\ \max=y_2=\frac{kN}{\sigma\sqrt{2\pi}}.e^0=\frac{kN}{\sigma\sqrt{2\pi}}$$

$$\therefore\ \frac{y_2}{y_1}=\frac{kN/\sigma\sqrt{2\pi}}{N/\sigma\sqrt{2\pi}}=k \text{ or } y_1=\frac{1}{k}y_2$$

i.e., the maximum frequency of the first is $\frac{1}{k}$ times the maximum frequency the second.

PROBLEM 13.32. *Show that the ordinates y (0), y (σ), y (2σ), y (3σ) at x=0, σ, 2σ, 3σ are* $\frac{1}{\sigma\sqrt{2\pi}}\left(1\ \frac{1}{\sqrt{e}},\ \frac{1}{e^2\sqrt{e}},\ \frac{1}{e^4\sqrt{e}}\right)$

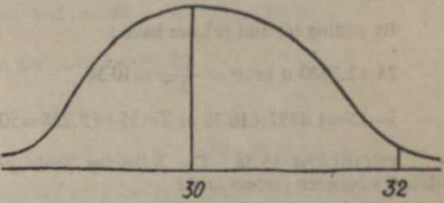
SOLUTION. The normal distribution is given by :

$$y=\frac{1}{\sigma\sqrt{2\pi}}e\text{-}\frac{x^2}{2\sigma^2}$$

x	ordinate	
$x=0$	$y(0)$	$=\frac{1}{\sigma\sqrt{2\pi}}e^0=\frac{1}{\sigma\sqrt{2\pi}}$
$x=\sigma$	$y(\sigma)$	$=\frac{1}{\sigma\sqrt{2\pi}}e\text{-}\frac{\sigma^2}{2\sigma^2}=\frac{1}{\sigma\sqrt{2\pi}}$
$x=2\sigma$	$y(2\sigma)$	$=\frac{1}{\sigma\sqrt{2\sigma}}e\text{-}\frac{4\sigma^2}{2\sigma^2}=\frac{e^{-2}}{\sigma\sqrt{2\pi}}=\frac{1}{e^2\sigma\sqrt{2\pi}}$
$x=3\sigma$	$y(3\sigma)$	$=\sigma\frac{1}{\sqrt{2\pi}}e\text{-}\frac{9\sigma^2}{2\sigma^2}=\sigma\frac{e^{-4.5}}{\sqrt{2\pi}}=\frac{1}{e\sigma^4\sqrt{2\pi}}$

PROBLEM 13.33. *A hundred squash balls are tested by dropping from a height of 100 centimetres and measuring the height of bounce. A ball is 'fast' if it rises above 32 centimetres. The average height of bounce was 30 centimetres and the standard deviation was $\frac{3}{4}$ centimetres. What is the chance of getting a 'fast' standard ball ?*

SOLUTION. If the mean is 30 and the standard deviation is $\frac{3}{4}$ the probability of getting a fast ball is given by the probability of the variate falling to the right of the ordinate $\frac{32-30}{3/4}=\frac{8}{3}$ $=2.67$.



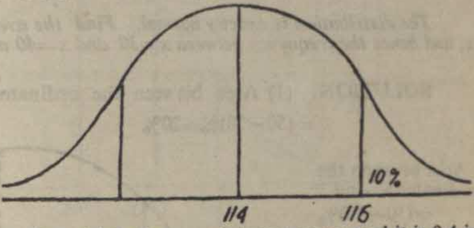
From the table the area between the mean and $x=2.67$ and standardised normal distribution is 0.4962.

$\therefore$ The probability that the value will be greater than $32=0.5-0.4962=0.0038$.

PROBLEM 13.34. *Sacks of grain packed by an automatic machine loader have an average weight of 114 kgs. It is found that 10% of the bags are over 116 kgs. Find the standard deviation.*

SOLUTION. The proportion of bags over 116 kgs. is 10%

$\therefore$ The proportion of bags between 114 kgs. and 116 kgs.$=40\%=0.4$



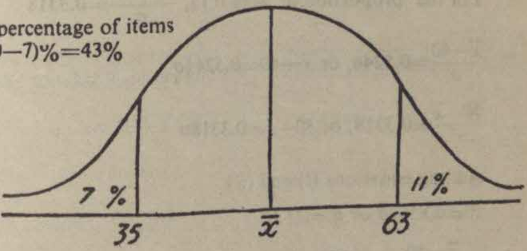
From the table the ordinate for which the area between mean and it is 0.4 is 1.2837.

$$\therefore \quad \frac{116-114}{\sigma}=1.2837 \text{ or } \sigma=\frac{2}{1.2837}=1.56$$

PROBLEM 13.35. *In a distribution, exactly normal, 7% of the items are under 35 and 89% are under 63. What are the mean and standard deviation of the distribution ?*

SOLUTION. The percentage of items between $x=35$ and $x=(50-7)\%=43\%$

The percentage of items between $x=63$ and $x=(89-50)\%$ $=39\%$



From the table

When the percentage of items is 43, the ordinate is 1.4757

When the percentage of items is 39, the ordinate is 1.2263

Therefore, we have

$\frac{\bar{x}-35}{\sigma}=1.4757$, or $\bar{x}-35=1.4757\sigma$ (i)

$\frac{63-\bar{x}}{\sigma}=1.2263$, or $63-\bar{x}=1.2263\sigma$ (ii)

By adding (*i*) and (*ii*), we have :

$28=2.7020\,\sigma$ or $\sigma=\frac{28}{2.7}=10.36$

$\bar{x}-35=1.4757\times 10.36$ or $\bar{x}=35+15.288=50.3$

PROBLEM 13.36. *The following table gives frequencies of occurrence of a variate x between certain limits :*

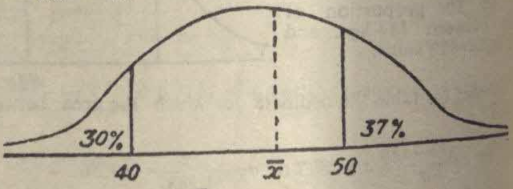
Variate (x)	*Frequency*
Less than 40	30
40 or more but less than 50	33
50 and more	37
Total	100

The distribution is exactly normal. Find the average and standard deviation of x, and hence the frequency between x=30 and x=40 and between x=50 and x=60.

SOLUTION. (1) Area between the ordinates 40 and $\bar{x}$ $=(50-30)\%=20\%$

Area between the ordinates 50 and $\bar{x}$ $=(50-37)\%$ $=13\%$

From the table



For the proportion of area 0.20, $\frac{x-\bar{x}}{\sigma}=0.5244$

For the proportion of area 0.13, $\frac{x-\bar{x}}{\sigma}=0.3318$

$\frac{\bar{x}-40}{\sigma}=0.5244$, or $\bar{x}-40=0.5244\sigma$ (i)

$\frac{50-\bar{x}}{\sigma}=0.3318$, or $50-\bar{x}=0.3318\sigma$ (ii)

Adding equations (*i*) and (*ii*)

$10=0.8562\sigma$ or $\sigma=11.68$

and $\frac{\bar{x}-40}{11.68}=0.5244$ or $\bar{x}=40+6.12=46.1$

(*ii*) The normal distribution has 46.1 as mean and 11.68 a standard deviation. Now the frequency between the mean and ordinate 40 is 20% and the frequency between mean $\frac{46.1-30}{11.68}=1.38$ and mean is equal to 41.62%—From table.

$\therefore$ Frequency between $x=30$ and $x=40$ is equal to $(41.62-20)\%$
$=21.62\%$

Frequency between the mean and $x=50=13\%$

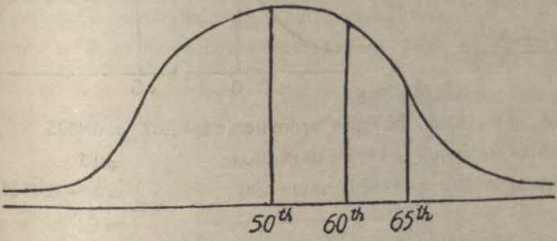
and $\frac{60-46.1}{11.68}=1.19$ is 38.3%—From table

$\therefore$ frequency between $x=50$ and $x=60=(38.3-13)\%=25.3$

PROBLEM 13.37. *The mean of a normal distribution is 50 and the standard deviation is 12.5. Calculate the 60th, 65th, 85th, 90th and 95th percentiles.*

SOLUTION. We know that 99 percentiles divide the area into 100 equal parts.

The standardised normal distribution will have the variate $\frac{x-50}{12.5}$



Now, in the adjoining diagram :

Area between the 50th and 60th percentile $=0.10$
Area between the 50th and 65th percentile $=0.15$
Area between the 50th and 85th percentile $=0.35$
Area between the 50th and 90th percentile $=0.40$
Area between the 50th and 95th percentile $=0.45$

From the table we find :

the standardised deviate for which the area is 0.1026 is 0.26 and for which the area is 0.0987 is 0.25 .:. the standardised deviate for which the area is 0.1 is 0.2545

Now, $\frac{x-50}{12.5}=0.2543$ or $x=50+3.178$

The 60th percentile is 53.178 (approx).

Similarly the 65th percentile is $\frac{x-50}{12.5}=0.385$

or $x=50+4.812=54.812$ (approx).

The 85th percentile is $\frac{x-50}{12.5}=1.04$

or $x=50+12.950=62.950$ (approx).

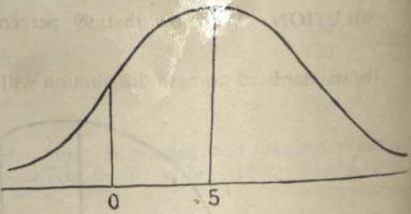
The 90th percentile : $\frac{x-50}{12.5}=1.29$ $\therefore$ $x=50+16.12=66.12$ (approx.)

The 95th percentile : $\frac{x-50}{12.5}=1.645$ or $x=50+20.32=70.32$ (approx.)

PROBLEM 13.38. *A normal distribution has mean 5 and standard deviation 3. What is the probability that the deviation from the mean of an item taken at random will have a negative value ?*

SOLUTION. The normal distribution is given by

$$P\ (x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{(x-\bar{x})^2}{2\sigma^2}}$$



Deviation from the mean will have a negative value when the number lies to the left of the ordinate at zero.

Here, $a=3$ and $\bar{x}=5$ and $x=0$

$$\frac{x-\bar{x}}{\sigma}=\frac{-5}{3}=-1.67$$

Area between the mean ordinate and—1.67 $=0.4525$

Area to the left of the mean ordinate $=0.5$

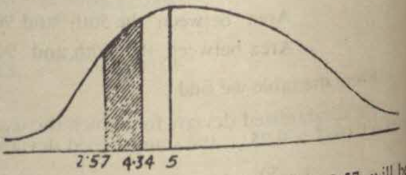
Area to the left of the ordinate at 0 $=0.5-0.4525$

$=.0475$

Therefore, the probability that the deviation from the mean of an item taken at random will have a negative value$=0.0475$.

PROBLEM 13.39. *Find the probability that an item drawn at random from the normal distribution with mean 5 and standard deviation 3 will be between*

(*i*) *2.75 and 4.34 ;*
(*ii*) *−1.24 and 1.37,*
(*iii*) *3.74 and 6.8.*



SOLUTION. (*i*) The probability is given by the shaded area in the adjoining figure.

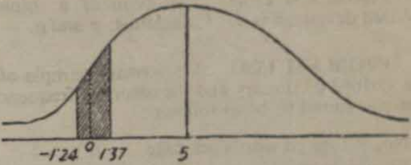
In the standardised normal distribution with mean 0, the deviate 2.57 will be given by :

$$\frac{x-\bar{x}}{\sigma}=\frac{2.57-5}{3}=\frac{-2.43}{3}=-0.81$$

and the deviate 4.34 will be given by :

$$\frac{x-\bar{x}}{\sigma}=\frac{4.34-5}{3}=\frac{-0.66}{3}=-0.22$$

Proportion of area between 0 and $-0.81 = 0.2910$ —from table
Proportion of area between 0 and $-0.22 = 0.0871$ —from table
Proportion of area between -0.81 and $-0.22 = 0.2910 - 0.0871 = 0.2039$



$\therefore$ Required Probability

$= 0.2039$

(*ii*) The probability is given by the shaded area in the adjoining figure.

In the standardised normal distribution, the deviate -1.24 will be given by

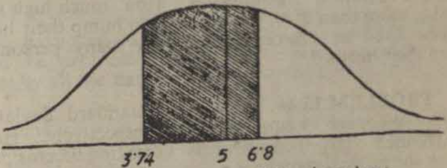
$$\frac{x-\bar{x}}{\sigma} = \frac{-1.24-5}{3} = \frac{-6.24}{3} = 2.08$$

and the deviate 1.37 will be given by

$$\frac{x-\bar{x}}{\sigma} = \frac{1.37-5}{3} = \frac{-3.63}{3} = -1.21$$

The proportion of area between o and $-2.08 = 0.4812$ } From table
The proportion of area between o and $-1.21 = 0.3869$

$\therefore$ The proportion of area between -2.08 and $-1.21 = 0.4812 - 0.3869$
$= 0.0943$



$\therefore$ Required Probability
$= 0.0943$

(*iii*) The probability is given by the shaded area in the figure.

In the standardised normal distribution, the deviate 3.74 is given by :

$$\frac{x-\bar{x}}{\sigma} = \frac{3.74-5}{3} = \frac{-1.26}{3} = -0.42$$

and the deviate 6.8 is given by

$$\frac{x-\bar{x}}{\sigma} = \frac{6.8-5}{3} = \frac{-1.8}{3} = 0.6$$

Proportion of area between o and $-0.42 = 0.1628$ } From table
Proportion of area between o and $0.\ 6 = 0.2257$
$\therefore$ Proportion of area between -0.42 and $0.\ 6 = 0.1628 + 0{\cdot}2257$
$= 0.385$

The probability $= 0.3885$.

REVISIONAL PROBLEMS—13

PROBLEM 13.40. Eight coins are thrown simultaneously. Find the chance of obtaining—(*i*) at least 6 heads (*ii*) no head (*iii*) all heads

$\left(\textit{Ans : (i)}\ \frac{37}{256}\ \textit{(ii)}\ \frac{1}{256}\ \textit{(iii)}\ \frac{1}{256} \right)$

PROBLEM 13.41. Take 100 sets of 10 tosses of an unbiased (perfect) coin. In how many cases do you expect to get (a) 7 heads and 3 tails, and (b) 7 heads at least.

(*Ans.* : (*i*) *12* (*ii*) *17*)

PROBLEM 13.42. The mean of a binomial distribution is 20, and the standard deviation is 4. Calculate *n*, *p and q*.

(*Ans.* : $n=100, p=\frac{1}{5}\ q=\frac{4}{5}$)

PROBLEM 13.43. A systematic sample of 100 pages was taken from the concise Oxford dictionary and the observed frequency distribution of foreign words per page was found to be as follows :

No. of foreign words per page	0	1	2	3	4	5	6
Frequency	48	27	12	7	4	1	1

Calculate the data by a Poisson distribution.

(*Ans.*	0	1	2	3	4	5	6)
	37.14	36.77	18.2	6.0	1.5	0.30	0.05

PROBLEM 13.44 In a certain factory turning out razor blades there is a small chance $\frac{1}{500}$ for any blade to be defective. The blades are in packet of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 1000 packets.

(*Ans* : $N(po)=9802$; $N(p_1)=196$; $N(p_2)=2$)

PROBLEM 13.45. Suppose that a doorway being constructed is to be used by a class of people whose heights are normally distributed with mean 170 centimetres and standard deviation 3 centimetres. How much high the doorway should be without causing more than 25% of the people to bump their heads. If the height of the door may be fixed at 176 centimetres, how many persons out of 5,000 are expected to bump their heads ?

(*Ans.* : *167.5* ; *2386 persons*)

PROBLEM 13.46. Mean and standard deviation of chest measurements of 1,200 soldiers are 85 cms. and 5 cms. respectively. How many of them are expected to have their chest measurements exceeding 95 cms., assuming the measurements to follow the normal pattern ?

(*Ans.* : *27*)

PROBLEM 13.47. The scores made by candidates in a certain test are normally distributed with mean 500 and standard deviation 100. What percent of candidates receive scores (i) less than 400, (ii) between 400 and 600 ? (For a standard normal distribution $Z=\frac{x-\bar{x}}{\sigma}$ the area under the curve between $Z=0$ and $Z=1$ is 0.34134.)

(*Ans.* : (*i*) *15.87%* (*ii*) *68.26%*)

PROBLEM 13 48. Calculate the frequencies of the normal distribution which has the same mean, standard deviation and total frequency as the distribution given below for the intervals 60—65, 65—70 etc.

x :	60—	65—	70—	75—	80—	85	90—	95—
f :	3	21	150	335	326	135	26	4

(*Ans.* : *3.1* ; *30.8* ; *148.0* ; *322.2* ; *319.5* : *144.0* ; *29.6* ; *2.8*)

PROBLEM 13.49. If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 both chosen at random (a) 1, (b) 0, (c) at most 2 bolts will be defective.

(*Ans.* : (*a*) *.4096* (*b*) *.4096* (*c*) *.9728*)

PROBLEM 13.50. The probability of a bomb hitting a target is §. Two bombs are enough to destroy a bridge. If six bombs are aimed at the bridge, find the probability that the bridge is destroyed.

(*Ans.* : *0.351*)

Sampling of Attributes—Large Samples

14

In dealing with problems of sampling we have chosen to distinguish between Attributes and Variables, and between large samples and small samples. Besides these we have presumed that in practically all these problems conditions of Simple Sampling hold good. By simple sampling is meant such random sampling in which each event has an equal chance of success and the chances of success of various events are independent, and not affected by previous trials.

In this chapter two types of problems have been solved.

(1) *Calculation of Standard Deviation.* Two formulae have been used for this purpose. For calculating σ of numbers, the formula used is $\sigma = \sqrt{npq}$ and for calculating σ of proportion, the formula used is

$$\sigma = \sqrt{\frac{pq}{n}}$$

By these two methods we have studied the significance of the difference between theory and practice and the problem of setting limits in the universe. It has been presumed that practically all the cases in a normal distribution would be between mean $\pm 3\sigma$.

(2) *Calculation of the standard error of the difference between proportions.* Standard error of the difference between proportions is calculated to find out if the difference between p_1, *i.e.*, the proportion of success in one sample and p_2, *i.e.* the proportion of success in another sample is significant or not. Here, it has also been presumed that if the difference is more than 3 times its standard error it is significant otherwise it could have arisen due to fluctuations of sampling.

Here we have made use of the following three types of formulae :—

(*a*) In those problems in which the presumption is that the proportion of A's is the material from which two samples have been drawn and, *i.e.* $p_1 = p_2$ the standard error of the difference or

$$s.e._{1-2} = \sqrt{p_0\, q_0 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}$$

(*b*) If p_1 is not equal to p_2 and p_1 and p_2 represent the true value of the proportion of A's in the two samples respectively then the

$$s.e._{1-2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

(*c*) If p_1 and p_2 are not equal but represent the true values of A's in the two samples respectively and if p_1 is not compared with p_2 but p_0, *i.e.*, the proportion of A's in the two samples together then

$$s.e._{1-2} = \sqrt{\frac{p_0 q_0}{n_1 + n_2} \times \frac{n_2}{n_1}}$$

STANDARD DEVIATION (OF NUMBERS)

PROBLEM 14.1. *A die is tossed 960 times and it falls with 6 upward 184 times. Is the die biased ?*

SOLUTION. On the null hypothesis the chance of getting a $6 = \frac{1}{6}$

Therefore $np = 960 \times \frac{1}{8} = 160$.
and the standard error or

$$\sigma = \sqrt{npq} = \sqrt{960 \times \frac{1}{6} \times \frac{5}{6}} = 11.54$$

The actual difference between observed and expected results is (184—160) =24, and it is 2.1 times the standard error. Hence, theoretically such a difference could have arisen due to sampling fluctuations as it is less than 3 times the standard error.

At 5% level of significance the hypothesis would have to be rejected because the difference is more than 1.96 times the standard error, and we shall have to conclude that the die is biased.

Yet another way of looking at the problem is that the probability of getting such a result due to chance is .03 or 3%. Hence, at 5% level the hypothesis has to be rejected and the conclusion drawn that the die is biased. At 1% or 2% level of significance the conclusion would be reverse. Usually results are tested at 5% level of significance.

PROBLEM 14.2. *A coin is tossed 400 times and it turns up head 216 times. Discuss whether the coin may be an unbiased one, and explain briefly the theoretical principles you would use for this purpose.*

SOLUTION. If the coin is unbiased, the chance that it will turn up head in one toss is $\frac{1}{2}$.

Therefore the expected number of heads in 400 tosses is 200. The observed number of heads is 216, and thus it is in excess of the expected number by 16. The standard deviation of simple sampling is

$$\sigma = \sqrt{npq} = \sqrt{\tfrac{1}{2} \times \tfrac{1}{2} \times 400} = 10$$

Where n represents the number of observations and p and q the probabilities of happening and not happening of an event respectively.

The deviation observed is thus only 1.6 times this standard deviation and it is therefore probable that it arose as a sampling fluctuation. Therefore, the coin can be said to be an unbiased one.

PROBLEM 14.3. *In tossing a hundred cents a student gets 66 heads. Do you think that he has used sufficient care to obtain a random toss each time ?*

SOLUTION. In a random toss, the chance of a student turning up head is $\frac{1}{2}$.

So, in a toss of 100 cents, the expected number of heads is 50.

The observed number of heads is 66, and thus it is in excess of the expectation by 16.

The standard deviation of simple sampling is

$$\sigma = \sqrt{npq} = \sqrt{\tfrac{1}{2} \times \tfrac{1}{2} \times 100} = 5.$$

The deviation observed is thus more than 3 times the standard error and it is most improbable that it arose as a sampling fluctuation. Therefore, it can be said that the student has not used sufficient care to obtain a random toss each time.

PROBLEM 14.4. *12 dice are thrown 3086 times and a throw of a 2, 3, 4 is reckoned as a success. Suppose that 19142 throws of 2, 3 and 4 have been made out. Do you think that this observed value deviates from the expected value ? If so, can the deviation from the expected value be due to fluctuations of simple sampling ?*

SOLUTION. The total number of throws $=3086\times12=37032$

The chance of success, that is, of throwing a

2, 3 and 4 with one dice in one throw $=\frac{1}{2}$.

Hence the expected value of successes $=\frac{1}{2}\times37032=18516$

The observed value of successes is 19142.

Thus the observed number of success is in excess of the expected number by $(19142-18516)=626$.

The standard deviation of sampling is

$$\sigma=\sqrt{npq}=\sqrt{\tfrac{1}{2}\times\tfrac{1}{2}\times37032}=96.3$$

Thus, deviation observed is 6.5 times this quantity and it is therefore most improbable that it is due to fluctuations of sampling.

PROBLEM 14.5. *Certain crosses of the pea gave 5321 yellow and 1804 green seeds. The expectation is 25 per cent of green seeds on a Mendelian hypothesis. Can the divergences from the expected values have arisen from fluctuations of simple sampling only ?*

SOLUTION. The total number of pea seeds examined $=(5321+1804)=7125$.

The expectation of green seeds is 25 per cent of the total.

$\therefore$ The expected result is 1781 green seeds. But the observed results is 1804 green seeds, and so it is in excess of the expected results by 23. The standard deviation of simple sampling is

$$\sigma=\sqrt{npq}=\sqrt{0.25\times0.75\times7125}=36.6$$

The divergence from expectation is thus only 0.6 times of this and hence may very well have arisen from fluctuations of simple sampling.

PROBLEM 14.6. *In breeding certain stocks, 408 hairy and 126 glabrous plants were obtained. If the expectation is one-fourth glabrous, is the divergence significant, or might it have occurred as a fluctuation of sampling ?*

SOLUTION. Total number of plants observed is $408+126=534$.

If the expectation of glabrous plants is one-fourth of the total, the expected number of glabrous plants is 133.5 or 134.

The observed number of glabrous plants is 126 which is less than the expected number by 8.

The standard deviation of simple sampling is

$$\sigma=\sqrt{npq}=\sqrt{\frac{1}{4}\times\frac{3}{4}\times534}=10 \text{ approx.}$$

Thus the divergence from the expectation is only .8 times of this and hence may very well have arisen from fluctuations of simple sampling.

PROBLEM 14.7. *Balls are drawn from a bag containing equal numbers of black and white balls each ball being returned before drawing another. In 2250 drawings 1018 black and 1232 white balls have been drawn. Do you suspect some bias on the part of the drawer ?*

SOLUTION. The expectation of drawing a white ball in one draw is $\frac{1}{2}$, since the bag contains equal number of black and white balls.

In 2250 drawings the expected number of white balls is 1125.

But the number of white balls drawn is 1232. Thus the numerical difference from the expected results is 107.

The standard deviation of simple sampling is

$$\sigma=\sqrt{npq}=\sqrt{\frac{1}{2}\times\frac{1}{2}\times2250}=23.7$$

The divergence from the expectation is thus about 4.5 times of this, hence it is not probable that it arose due to fluctuations of sampling.

Explanation of the deviation must be sought somewhere else, and it se reasonable to suspect that the drawer was biased.

PROBLEM 14.8. *In a sample of 100 in Central U.P., 60 are found wheat-eaters and 40 rice-eaters. Can we assume that both the food articles equally popular ?*

SOLUTION. Taking the hypothesis that both the food articles are equ popular, the expected frequency for wheat-eaters is 50.

The observed frequency in 60.
Thus the difference between the two is 10.
Standard deviation of sampling is.

$\sigma = \sqrt{npq} = \sqrt{\frac{1}{2} \times \frac{1}{2} \times 100} = 5.$

The divergence observed is thus only twice this standard error and he may have arisen from fluctuations of simple sampling. Therefore, the assumpt that both the food articles are equally popular cannot be a mistaken notion.

STANDARD DEVIATION (OF PROPORTIONS)

PROBLEM 14.9. *A group of scientific men reported 1,705 sons and 1,5 daughters. Do these figures conform to the hypothesis that the sex ratio is ½?*

SOLUTION. The total number of observation is 1705 + 1527 = 3232.

The number of sons is 1705.
Therefore the observed male ratio is 1705/3232 or 0.5275.
On the given hypothesis the male ratio ough t to be 0.5000.
Thus the observed male ratio is in excess of the theoretical ratio by 0.0275
The standard deviation of the proportion is

$$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{3232}} = .0088$$

The divergence from hypothesis is thus about 3.13 times of this standa error and it is therefore most improbable that it arose as a sampling fluctuatio Hence it can be definitely said that the figures given do not conform to the giv hypothesis.

PROBLEM 14.10. *In a group of 50 first cousins there were found to be 2 males and 23 females. Ascertain if the observed proportions are inconsistent wi the hypothesis that the sexes should be in equal proportion.*

SOLUTION. The number of cousins observed is 50 and the number male cousins is 27.

Therefore the proportion of male cousins is $\frac{27}{50}$ or 0.54.
If the hypothesis is true it ought to be 0.50.
Thus the observed deviation is .04.
The standard deviation of the proportion is

$$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{50}} = .07.$$

The divergence from the hypothesis is only ·57 of this standard error Hence it can be said that the observed proportions are not inconsistent with th given hypothesis.

PROBLEM 14.11. *12 dice were thrown 6500 times, 4, 5 and 6 being recko- ned as a "success." What proportion of success do you expect ? If in actual observation, the proportion of success is found to be 0.5016, find the standard devia- tion of the proportion with the given number of throws and state whether you would regard the excess of success as probably significant of bias in the dice.*

SOLUTION. The total number of throws $=6500\times 12=78000$.

The expected proportion of successes is $\frac{1}{2}$ or 0.50000.

The observed proportion of successes is 0.5016, and thus is in excess of expected proportion by .0016.

The standard deviation of the proportion is

$$\sigma=\sqrt{\frac{pq}{n}}=\sqrt{\frac{1}{2}\times\frac{1}{2}\times\frac{1}{78000}}=.00179$$

The deviation observed is only .9 times this quantity and it is, therefore, probable that it arose as a sampling fluctuation. Therefore, the excess of the proportion of successes is not significant of bias in the dice.

PROBLEM 14.12. *1000 ladies are chosen at random from the inhabitants of Maharashtra and it is found that 55 per cent of them have dark eyes and the remainder have eyes of some other colour. What can be inferred about the proportion of dark-eyed ladies in the whole population of Maharashtra ?*

SOLUTION. Assuming that the conditions of simple sampling hold, we have,

p, that is the chance of getting a dark-eyed lady $=0.55$

and q, the chance of failure $=0.45$

Total number of observations, or n $=1000$

Standard Error of the simple sampling is

$$\sigma=\sqrt{npq}=\sqrt{0.55\times 0.45\times 1000}=15.7$$

Whatever be the actual value of p, a simple sample should give a number of success within 3 times this standard error, *i.e.* $15.7\times 3=47.1$ of the expected frequency pn. This is 4.71 per cent of the sample.

Thus, it can be said that the percentage of dark-eyed ladies in the whole population is 55 ± 4.71, that is, it lies between 50.29 and 59.71 per cent.

PROBLEM 14.13. *In a locality containing 18,000 families, a sample of 840 families was selected at random. Of these 840 families, 206 families were found to have a monthly income of Rs. 250 or less. It is desired to estimate how many out of the 18,000 families have a monthly income of Rs. 250 or less. Within what limits would you place your estimate ?*

SOLUTION. Proportion of families having a monthly income of Rs. 250 or less, *i.e.*, p is $\frac{206}{840}=\frac{103}{420}$, and so q, *i.e.* the proportion of families not having a monthly income of Rs. 250 or less, is $\frac{317}{420}$. Assuming that the conditions of sampling hold true.

Standard error of the proportion of families having a monthly income of Rs. 250 or less is

$$\sigma=\sqrt{\frac{pq}{n}}=\sqrt{\frac{103}{420}\times\frac{317}{420}\times\frac{1}{840}}$$

$$=0.015=1.5 \text{ per cent.}$$

Whatever be the percentage of families who have a monthly income of Rs. 250 or less, a simple sample should give a percentage within three times this standard error.

Hence taking $\frac{103}{420}$ (or 24.5%) to be the estimate of families having a monthly income of Rs. 250 or less, in the locality, we have that the limits are 24.5 per cent, $\pm$ 4.5 per cent, *i. e.* 20 per cent and 29 per cent approximately.

PROBLEM 14.14. *400 children are chosen in an industrial town of Northern India, and 150 are found to be under weight. Assuming the conditions of simple sampling, estimate the percentage of children who are under weight in the industrial town and assign limits within which the percentage probably lies.*

SOLUTION. Taking the observed values of p and q in the sample we have

p, that is the chance of getting an under weight child $= \frac{150}{400} = \frac{3}{8}$
and q, that is the chance of failure $= 5/8$
Total number of children examined, or n is 400.
Substituting the above values of p, q and n in the given formula, we get,
Standard error of the proportion of children who are underweight

$$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{3}{8} \times \frac{5}{8} \times \frac{1}{400}} = 0.024 = 2.4 \text{ per cent.}$$

Whatever may be the percentage of children who are under weight in the population, a simple sample should give a percentage within three times this standard error.

Hence, taking 3/8(or 37.5%) to be the estimates of the number of children who are under weight, we have that the limits are 37.5 per cent ± (3×2.4) per cent that is 30.3 and 44.7 per cent approximately.

PROBLEM 14.15. *500 eggs are taken at random from a large consignment, and 50 are found to be bad. Estimate the percentage of bad eggs in the consignment and assign limits within which the percentage probably lies.*

SOLUTION. Proportion of bad eggs in the sample or $p = 1/10$
and so q $= 9/10$
Standard error of the proportion of bad eggs

$$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{1}{10} \times \frac{9}{10} \times \frac{1}{500}} = 0.013 = 1.3 \text{ per cent.}$$

Whatever may be the percentage of bad eggs in the consignment, a simple sample should give a percentage within three times this standard error.

Hence taking 1/10(or 10%) to be the estimate of bad eggs we have that the limits are 10 per cent ± 3.9 per cent, that is, 6.1 per cent and 13.9 per cent approximately.

PROBLEM 14.16. *Explain the terms (a) Statistic (b) Parameter (c) Standard error of a statistic.*

A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Estimage the proportion of the bad pineapples in the consignment, as well as the standard error of the estimate. Deduce that the percentage of bad pineapples in the consignment almost certainly lies between 8.5 and 17·5.

SOLUTION. Proportion of bad pineapples in the sample or

$$p = \frac{65}{500} \text{ and so } q = \left[1 - \frac{65}{500}\right] = \frac{435}{500}$$

Standard errors of the proportion of bad pineapples

$$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{65}{500} \times \frac{435}{500} \times \frac{1}{500}}$$

= .015 or 1.5 per cent.

The percentage of bad pineapples in the sample is $\left(\frac{65}{500} \times 100\right) = 13$ so the limits of bad pineapples in the consignment would be $\{13 \pm (3 \times 1.5)\}$ per cent or 8.5 per cent and 17.5 per cent.

PROBLEM 14.17. *A sample of 1000 days is taken from meteorological records of a certain district, and 120 of them are found to be foggy. What are the probable limits to the percentage of foggy days in the district ?*

SOLUTION. *Proportion of foggy days or*

$$p=\frac{120}{1000}=\frac{3}{25} \text{ and } q=\frac{22}{25}$$

Standard error of the proportion of foggy days

$$\sigma=\sqrt{\frac{pq}{n}}=\sqrt{\frac{3}{25}\times\frac{22}{25}\times\frac{1}{1000}}$$

$=0.0103=1.03$ per cent

Whatever may be the percentage of foggy days in the whole district, a simple sample should give a percentage within three times the standard error.

Hence taking 3/25 (or 12%) to be the estimate of foggy days we have that the limits are 12 per cent ± 3.09 per cent, that is 8.91 per cent and 15.09 per cent approximately.

PROBLEM 14.18. *In a newspaper article of 1600 words in Hindustani 64 per cent of the words are found to be of Hindi origin, assuming the simple sampling conditions hold good, estimate the proportion of Hindi words in the writer's vocabulary and assign limits to that proportion.*

SOLUTION. Assuming the conditions of simple sampling to hold true, we have,

Proportion of words of Hindi origin or $p=\frac{64}{100}$ and $q=\frac{36}{100}$. Standard error of the proportion of words of Hindi origin

$$\sigma=\sqrt{\frac{pq}{n}}=\sqrt{\frac{64}{100}\times\frac{36}{100}\times\frac{1}{1600}}=0.012=1.2 \text{ per cent.}$$

Whatever may be the percentage of words of Hindi origin in the writer's vocabulary, a simple sample should give a percentage within three times this standard error.

Hence taking 64/100 or (64%) to be the estimate of words of Hindi origin in the writer's vocabulary, we have that the limits are 64 per cent ± 3.6 per cent., *i.e.*, 60.4 per cent and 67.6 per cent approximately.

PROBLEM 14.19. *A life insurance company founded upon the Indian Experience Table has a thousand policies averaging Rs. 2,000 on lives at age 25. From the experience table it is found that of 89,032 alive at age 25, 88,000 are alive at age 26. Find the upper and lower values of the amount that the company will have to pay out in insurance during the year.*

SOLUTION. The chance that a man of age 25 would not be alive at age 26 is

$$p=1-\frac{88000}{89032}=\frac{1032}{89032}$$

Therefore, the expected number of deaths out of 1000 insured persons is

$$=\frac{1000\cdot 1032}{89032}=12 \text{ approx.}$$

The standard error of the simple sampling is

$$\sigma=\sqrt{npq}=\sqrt{\frac{1032}{89032}\times\frac{88000}{89032}\times1000}=3.4.$$

The actual number of deaths would lie between the expected number of deaths ± three times this standard error, *i.e.* $12\pm3\times(3.4)$, or 1.8 and 22.2.

The average amount of a policy is 2,000 rupees. Therefore, the upper and lower values of the amount that the company will have to pay out in insurance during the year are Rs. $2{,}000 \times 1.8$ and Rs. $2{,}000 \times 22.2$ or Rs. 3,600 and 44,400.

PROBLEM 14.20. *A dealer takes 100 samples from a consignment of 10,000 items of a certain goods and finds that there are 50 items of grade I worth Rs. 50 per thousand, 30 items of grade II worth Rs. 40 per thousand, and 20 items of grade III worth Rs. 30 per thousand. Within what limits should the value of the consignment be fixed ?*

SOLUTION. Obviously, the true value of the consignment must be somewhere between Rs. 50 and Rs. 300 since all the items are neither of grade I nor of grade III. In order further to limit the values, the probable percentage of each grade in the whole shipment on the basis of the sample is found out.

Chance of getting Grade I Commodity is $p_1=\frac{1}{2}=0.5$
and $q=1-p=1-\frac{1}{2}=\frac{1}{2}=0.5$
Chance of getting Grade II Commodity is $p_1=0.3$
and so $q_1=0.7$
Similarly chance of getting Grade III Commodity is $p_1=0.2$
and so $q_1=0.8$
Total number of observations in the sample $=100$.
The standard error of the simple sampling is

$\sigma=\sqrt{npq}$

Substituting the value of n, p, and q, we get
σ for Grade I $=\sqrt{0.5\times0.5\times100}=5.0$
σ for Grade II $=\sqrt{0.3\times0.7=100}=4.6$
σ for Grade III $=\sqrt{0.2\times0.8\times100}=4.0$

Whatever be the true *pi*, *pii* and *piii* in the consignment, a simple sample should give a percentage within three times this standard error.

Hence, adding to and subtracting from the sample percentage, three times respective standard error for each grade, the following upper and lower value for three kinds of goods are found :

	Grade I [50±3 (5.0)]	Grade II [30±3 (4.6)]	Grade III [20±3 (4.0)]
Upper value	65%	43.8%	32%
Lower value	35%	16.2%	8%

Thus the highest value that can be placed upon the consignment is that value for which grade I is the highest and grade III the lowest that is,

Grade I=65%
Grade III 8%
73%
Grade II =27%

From the above figures it is found that the maximum value of the consignment is

Rs. 50 (65% of grade I)+Rs. 40 (27% of grade II)
+Rs. 30 (8% of grade III)=Rs. 325.0+Rs. 108.0+Rs. 24≃Rs. 457

Similarly, the lowest value that can be set on the consignment is that value for which grade I is the lowest and grade III the highest, that is

Grade I = 35% Grade III = 32% = 67% Grade II = 33%

From these figures, it is found that the minimum value of the shipment is Rs. 50 (35% grade I) + Rs. 40 (33% of grade II) + Rs. 30 (32% of grade III) = Rs. 175 + Rs. 132 + Rs. 96 = Rs. 403

The value of shipment can thus be fixed within the limits Rs. 403 and Rs. 457.

PROBLEM 14.21. *A man buys 1,000 sacks of potatoes. He finds that from 1,000 potatoes chosen from the sacks at random, 400 are of class A worth Rs. 100 a sack; 250 are of class B, worth Rs. 70 per sack; 200 are of class C, worth Rs. 50 a sack and 150 are of class D, worth Rs. 40 per sack. What are the upper and lower bounds for the values of the potatoes ?*

SOLUTION. In order to find the upper and lower bounds of the potatoes, the probable limits are first assigned to each grade in the universe.

Chance of getting class A or $p_a = 0.4$ so $q_a = 0.6$;

Chance of getting class B or $p_b = 0.25$ so $q_b = 0.75$

Chance of getting class C or $p_c = 0.20$ so $q_c = 0.80$

Chance of getting class D or $p_d = 0.15$ so $q_d = 0.85$

The standard error of the simple sampling is

$$\sigma = \sqrt{npq}$$

Substituting the respective values of n, p and q

We get.

σ for class A $= \sqrt{0.4 \times 0.6 \times 1000} = 15.5 = 1.55$ per cent

σ for class B $= \sqrt{0.25 \times 0.75 \times 1000} = 13.7 = 1.37$ per cent

σ for class C $= \sqrt{0.20 \times 0.80 \times 1000} = 12.65 = 1.265$ per cent

σ for class D $= \sqrt{0.15 \times 0.85 \times 1000} = 11.3 = 1.13$ per cent

Whatever be the true p_a, p_b, p_c, and p_d in the lot a simple sample should give a percentage within three times this standard error.

Hence adding to and subtracting from the sample percentages three times their respective standard errors.

The following upper and lower values for the four kinds of potatoes are found.

	Class A [40±3(1.55)]	Class B [25±3(1.37)]	Class C [20±3(1.265)]	Class D [15±3 (1.13)]
Upper value	44.65%	29.11%	23.795	18.39%
Lower value	35.35%	20.89%	16.205	11.61%

Thus the highest value that can be placed upon the whole lot is that value for which class A is the highest and classes C and D the lowest,

that is

Class A 44.65%
Class C 16.205%
Class D 11.61%

72.465%

Class B 27.535%

From the above figures, it is found that the maximum value of the lot is

Rs. 100(44.65% of class A) + Rs. 70(27.535% of class B)
+ Rs. 50(16.205% of class C) + Rs. 40(11.61% of class D)
= Rs. 44650 + Rs. 19274.5 + Rs. 8102.5 + Rs. 4644
= Rs. 76671.

Similarly the lowest value that can be set on the lot is that value for which classes C and D are the highest and class A the lowest,

that is

Class A	35.35 %
Class C	23.795%
Class D	18.39 %
	77.535%
Class B	22.465%

From those figures, it is found that the minimum value of the shipment is

Rs. 100(35.35% of class A)+Rs. 70(22.465% of class B)
+Rs. 50(23.795% of class C)+Rs. 40(18.39% of class D)
=Rs. 35350+Rs. 15725.5+Rs. 11897.5+Rs.7356
=Rs. 70329.

Thus the upper and the lower bounds for the value of potatoes is Rs. 76671 and Rs. 70329.

STANDARD ERROR (OR THE DIFFERENCE)

PROBLEM 14.22. *In a random sample of 1000 persons from town A, 400 are found to be consumers of rice. In a sample of 800 from town B, 400 are found to be consumers of rice. Discuss the question whether the data reveal a significance between A and B so far as the proportion of rice consumers is concerned.*

SOLUTION. In the two towns together, the percentage of rice consumers is

$$p_0=\frac{(400+400)\times 100}{1000+800}=44.4$$

and therefore

$$q_0=(100-44.4)=55.6$$

In town A it is 40 per cent whereas in town B it is 50 per cent. The difference in the percentages of rice consumers in the two towns is 10. Assuming that the samples taken are simple samples, then the standard error of sampling for the difference between percentages observed in the samples of the given sizes would be :

$$s.e._{1\text{-}2}=\sqrt{p_0 q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Where $s.e._{1\text{-}2}$ represents the standard error of the difference between two percentages or proportions and p_0 and q_0 the presence and absence of an attribute and n_1 and n_2 the number of items in the two samples respectively.

$$\sqrt{(44.4\times 55.6)\left(\frac{1}{1000}+\frac{1}{800}\right)}$$

$$=2.357 \text{ per cent.}$$

The actual difference which is 10 per cent, is over 4.2 times this standard error. So it can be concluded that the data reveal significant difference between A and B so far as the proportion of the consumers is concerned .

PROBLEM 14.23. *Show how you would test significance of the difference between the prevalence of a certain attribute in two given populations from each of which you could take large samples ?*

In a random sample of 500 men from a particular district of U. P.. 300 are found to be smokers. In one of 1,000 men from another district, 550 are smokers. Do the data indicate that the two districts are significantly different with respect to the prevalence of smoking among men ?

SOLUTION. In the two districts together, the percentage of smokers is :

$$p_0 = \frac{(300+550)\times 100}{500+1000} = 56.7\%$$

and therefore

$$q_0 = (100-56.7) = 43.3.$$

In first district the percentage of smokers is 60% whereas in other district is 55%. The difference in the percentages of smokers in the two districts is 5%. Assuming that the samples taken are simple samples, the standard error of sampling for the difference between percentages observed in the samples of the given sizes would be

$$s.e._{1-2} = \sqrt{p_0 q_0 \left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{(56.7\times 43.3)\left(\frac{1}{500}+\frac{1}{1000}\right)}$$

$$= 2.714 \text{ per cent.}$$

The actual difference which is 5 per cent, is 1.84 times the standard error. So it can be concluded that the data reveal an insignificant difference between the two districts of U. P., so far as the prevalence of smoking is concerned.

PROBLEM 14.24. *The following table gives the proportion of dark-coloured people in two cities.*

City	*Total population observed*	*Percentage of dark-coloured*
A	*250*	*42*
B	*450*	*33*

Can the difference observed in the percentage of dark-coloured people be due to the fluctuations of sampling ?

SOLUTION. In the two cities together the percentage of dark-coloured people is :

$$p_0 = \frac{(105+149)\times 100}{250+450} = 36.4 \text{ approx.}$$

and therefore $q_0 = (100-36.4) = 63.6$ approx.

If this were the true percentage, the standard error of sampling for the difference between percentages observed in samples of the given sizes would be

$$s.e._{1-2} = \sqrt{p_0 q_0 \left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$= \sqrt{(36.4\times 63.6)\left(\frac{1}{250}+\frac{1}{450}\right)} = 3.8 \text{ per cent approx.}$$

The actual diffeernce is 9 per cent and is 2.4 times that standard error, hence, this difference observed can be attributed to the fluctuations of sampling though it cannot be very definitely said so. Usually if the actual difference is more than twice the standard error of the difference it is considered to be significant.

PROBLEM 14.25. *In town A, 19400 persons were observed and 27% of them were found to be short-sighted. In town B, 29750 persons were observed and 40% were found to be short-sighted. Can the difference observed in the percentage of short-sighted persons be attributed solely to the fluctuations of sampling ?*

SOLUTION. In the two towns together the percentage of short-sighted persons is 28.8. If this were the true percentage, the standard error of sampling for the difference between percentages observed in the samples of the given sizes would be

$$s.e._{1\text{-}2}=\sqrt{p_0q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$=\sqrt{(28.8)\,(71.2)\left(\frac{1}{19400}+\frac{1}{29750}\right)}=0.42 \text{ per cent.}$$

The actual difference which is 3 per cent, is over 7 times this standard error and so cannot be attributed solely to the fluctuation of sampling.

PROBLEM 14.26. *One type of aircraft is found to develop engine trouble in 5 flights out of a total of 100 and another type in 7 flights out of a total of 200 flights. Is there a significant difference in the two types of aircrafts so far as engine defects are concerned ?*

SOLUTION. On the null hypothesis the proportion of engine defects to the total number of flights or

$$p_0=\frac{5+7}{100+200}=\frac{12}{300}=.04 \text{ and } q_0=(1-.04)=.96$$

The standard error of the difference in the proportions

$$s.e._{1\text{-}2}=\sqrt{p_0q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\sqrt{.04\times.96\left[\frac{1}{100}+\frac{1}{100}\right]}=.028$$

The actual difference in two proportions (.05 and .035) is .015. It is less than one standard error and as such the difference is not significant.

PROBLEM 14.27. *One thousand articles from a factory are examined and found to be three per cent defective. Fifteen hundred similar articles from a second factory are found to be only 2 per cent defective. Can it reasonably be concluded that the product of the first factory is inferior to the second ?*

SOLUTION. In the two factories together, the percentage of defective articles is

$$=\frac{(30+30)\times100}{1000+1500}=2.4$$

If this were true percentage, the standard error of sampling for the difference between percentages observed in samples of the given sizes would be

$$s.e._{1\text{-}2}=\sqrt{(p_0q_0)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$=\sqrt{2.4\times97.6\left(\frac{1}{1000}+\frac{1}{1500}\right)}=0.649 \text{ per cent approx.}$$

The actual difference is 1 per cent and is only 1.6 times this standard error and so could have arisen due to fluctuations of simple sampling,

Hence, it cannot be reasonably concluded that the product of the first factory is inferior to that of the second.

PROBLEM 14.28. *In 1970, for white males in the age group 30—34 years, the number dying in Chicago was 902 out of 106, 307 ; in New York 2130 out of 221,598 (United States Life Tables). Are the death rates in the two cities significantly different ?*

SOLUTION. In Chicago, the death rate is 8.48 per thousand.

In New York it is 9.61 per thousand. In the two cities together the death rate per thousand is

$$=\frac{(902+2130)\times 1000}{106307+221598}=9.25 \text{ approx.}$$

If this were the true death rate, the standard error of sampling for the difference in proportion observed in the samples of the given size under the conditions of simple sampling would be

$$s.e._{1\text{-}2}=\sqrt{p_0q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\text{or } = \sqrt{9.25\times 990.75\left(\frac{1}{106307}+\frac{1}{221598}\right)}=.1$$

The actual difference is 2.13 per thousand and the standard error of the difference is only .1. The actual difference is more than 10 times the standard error of the difference and so the death rates in the two cities are significantly different from each other.

PROBLEM 14.29. *In 1970, in the original registration states, the number of white males dying between the ages of 30 and 31 was 1609 out of a population of 253.445 of white males in this age group ; the corresponding figures for white females of the same age were 1302 out of 239,912. "United States Life Tables." Is there a significant difference between the death rates, of the two sexes at this age ?*

SOLUTION. The death rate per thousand for the two sexes in the given states for the age group 30—31 is

$$p_0=\frac{(1609+1302)\times 1000}{253445+239912}=4.9 \text{ approx.}$$

$$\therefore q_0=1000-p_0=1000-4.9=995.1$$

The death rate per thousand for males is

$=1609\times 1000/253445=6.3$ approx.

The death rate per thousand for females is

$=1302\times 1000/239912=5.4$ approx.

The difference between the two death rates is $6.3-5.4=$or 0.9. Assuming that the conditions of simple sampling hold true, the standard error of this difference is

$$s.e._{1\text{-}2}=\sqrt{p_0q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Substituting the values, we get

$$s.e._{1\text{-}2}=\sqrt{(5.9\times 995.1)\left(\frac{1}{253445}+\frac{1}{239912}\right)}=.22.$$

The observed difference is thus about 4 times this standard error. Hence it can be concluded that there is a significant difference between the death rates of the two sexes at this age.

PROBLEM 14.30. *In a certain association table, the following frequencies were obtained :*

$(AB)=927,\ (A\beta)=642\ ;\ (aB)=396\ (a\beta)=357.$

Can the association of the table have arisen as a fluctuation of simple sampling, the true association being zero.

SOLUTION. From the given association table

$$(A)=(AB)+(A\beta)=927+642=1569$$
$$(B)=(AB)+(aB)=927+396=1323=n_1$$
$$(\beta)=(A\beta)+(aB)=642+357=999=n_2$$
$$N=(B)+(\beta)=1323+999=2322$$

The percentage of A in B's is $=(AB)/(B)=927\times 100/1323=70.1$
The percentage of A in β's is $=(A\beta)/(\beta)=642\times 100/999=64.3$
The difference of the two percentages is $70.1-64.3$, or 5.8
The percentage of A in the universe or

$$p_0=\frac{(A)\times 100}{N}=\frac{1569\times 100}{2322}=67.6.$$

$$q_0=100-67.6=32.4.$$

Assuming that the conditions of simple sampling hold true, the standard error of the difference between the two percentages is

$$s.e._{1\text{-}2}=\sqrt{p_0q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Substituting the given values we get

$$s.e._{1\text{-}2}=\sqrt{(67.6\times 32.4)\left(\frac{1}{1323}+\left(\frac{1}{999}\right)\right.}$$

$=1.96$ per cent.

The observed difference which is 5.8 per cent is thus 2.96 times this standard error and so cannot be attributed to the fluctuations of simple sampling. Hence the association is a real one.

PROBLEM 14.31. *The figures for antitoxin treatment in the city of Toronto Hospital, for a certain period, in the treatment of diphtheria were:—*

	Cases	*Deaths*
Antitoxin treatment	*228*	*37*
Ordinary treatment	*337*	*28*

Can it be concluded that there were significantly more deaths in the group treated by antitoxin.

SOLUTION. On the null hypothesis the proportion of deaths for the whole group is :—

$$p_0=\frac{\text{total deaths}}{\text{total cases}}=\frac{65}{565}=.115.$$

$$=q_0=(1-.115)=.885.$$

The standard error of the difference in proportions between two samples is

$$s.e._{\ 1\text{-}2}=\sqrt{p_0q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$=\sqrt{.115\times .885\left(\frac{1}{228}+\frac{1}{337}\right)}=.027$$

The actual difference in the two proportions $\frac{37}{228}$ and $\frac{28}{337}$ (.162—.083) = .08 and it is above 2.8 times the standard error and as such very significant.

PROBLEM 14.32. *During a three-year period, experiments with two types of lighting systems were conducted in an elementary school. Room I had two 150-watt, manually controlled lights, which were turned on and off by teachers or pupils as needed. Room 2 had four 300 watt indirect lights, controlled by automatic relays which turned the lights on when additional illumination was needed and off when no longer required. The pupils were sixth grade students. All were given the "Otis self-administered Test of Mental Ability" and the "Standard Achievements Tests." They were divided equally between Room 1 and Room 2 according to the results of these tests. The two class-rooms were identical second floor rooms with north light. Two teachers employing the same teaching methods were assigned to departmental work with the two classes, each teacher teaching certain subjects in both sections. During the three-year period, there were 115 pupils in Room I and 112 pupils in Room 2. In Room I there were 29 failures and in Room 2 there were 9 failures. The percentages failing were 25.2 per cent for Room I, and 8.0 per cent for Room 2. Is there a significant difference between the two percentages ?*

SOLUTION. Assuming that the difference is a real one, the standard error of sampling for the difference between percentages observed in the two samples of the given sizes would be:

$$s.e.=\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$$

Substituting the given value, we get

$$s.e.=\sqrt{\frac{25.2\times 74.8}{115}+\frac{8\times 92}{112}}=4.8 \text{ per cent.}$$

The actual difference is 25.2—8 or 17.2 per cent which is 3.6 times this standard error and so could not have arisen through the chances of simple sampling. Therefore, the difference between the two percentages is significant.

Note. If we assume that the proportions of p's are not the same in the material from which the two samples are drawn, but p_1 and p_2 are the true values of the proportions, the standard error of sampling for the difference in the proportions of p in the two samples should be calculated by this formula.

PROBLEM 14.33. *The following table relates to the hair colour of girls at Bombay.*

	Of Dark Hair Colour	*Total observed*	*Per cent. Dark*
Bombay	*2,152*	*4,950*	*43.5*
Film Sector of Bombay	*400*	*974*	*41.1*
Non-Film Sector of Bombay	*1752*	*3,976*	*44.1*

Do you regard the difference observed in the percentages of girls of dark hair colour in Bombay and its Film Sector as significant ?

SOLUTION. The observed difference in the percentages of girls of dark hair colour in Bombay and its Film Sector is 2.4.

The standard error of sampling for the difference between these two percentages would be

$$s.e._{1-2}=\sqrt{\frac{p_0q_0}{n_1+n_2}\times\frac{n_2}{n_1}}$$

$$=\sqrt{\frac{43.5\times 56.5\times 39764}{49507\times 9743}}=0.45 \text{ per cent.}$$

The actual difference is over five times of this and could not have occu as a mere error of sampling Thus, the difference observed in the percentages girls of dark hair colour in Bombay and its Film Sector is significant.

Note :—When two samples are drawn from distinct materials or differ universes giving proportions of *A's* p_1, and p_2, but in lieu of comparing the p portion p_1, with p_2 it is compared with the proportion of A's in the two sampl together, *viz.*, p_0, this formula should be used.

PROBLEM 14.34. *The subject under investigation is the measure of dep dence of Tamil on words of Sanskrit origins. One newspaper articles reporting proceedings of the Constituent Assembly contained 2025 words of which 729 wor were declared by a literary critic to be of Sanskrit origin. A second article by same author describing atomic research contained 1600 words of which 640 wor were declared by the same critic to be of Sanskrit origin. Assuming that sim sampling conditions held estimate the limits for the proportion of Sanskrit terms the writer's vocabulary, and examine whether there is any significant difference the dependence of this writer on words of Sanskarit origin in writing on these tw subjects.*

SOLUTION. (*a*) The percentage of Sanskrit terms in the writer's vocabu lary in both the articles taken together is

$$p_0 = \frac{(729+640)\times 100}{2025+1600} = 37.77$$

$$\therefore q_0 = 100 - p_0 = 100 - 37.77 = 62.23.$$

The proportion of Sanskrit terms in the article reporting the proceedings o the C. A., is 729 × 100/2025 = 36 per cent. The proportion of Sanskrit terms i the article describing atomic research is 640 × 100/1600 = 40 per cent. Therefore the difference in the dependence of the writer on words of Sanskrit origin i writing or these two subjects is 40—36, or 4 per cent.

The standard error of sampling for the difference between these two per- centages is

$$s.e._{1\text{-}2} = \sqrt{p_0 q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Substituting the values, we get

$$s.e._{1\text{-}2} = \sqrt{(37.77\times 62.23)\left(\frac{1}{2025}+\frac{1}{1600}\right)} = 2.63 \text{ per cent.}$$

The observed difference is thus about 1.5 times this standard error and so could not have arisen through fluctuations of simple sampling.

Hence it can be concluded that there is a significant difference in the dependence of the writer on words of Sanskrit origin in writing on the two given subjects.

(*b*) Estimation of the limits for the proportion of Sanskrit terms in the writer's vocabulary.

Proportion of Sanskrit terms in the writer's vocabulary is

$$p = \frac{729+640}{2025+1600} = \frac{1369}{3625} \qquad \therefore q = 1 - \frac{1369}{3625} = \frac{2256}{3625}$$

Assuming that the conditions of simple sampling hold true,

The standard error of the proportion of Sanskrit terms in the writer's vocabulary is

$$\sigma = \sqrt{pq/n} = \sqrt{\frac{1369}{3625}\times\frac{2256}{3625}\times\frac{1}{3625}} = .0081 = .81 \text{ per cent.}$$

Whatever be the percentage of Sanskrit terms in the writer's vocabulary a simple sample should give a percentage within three times this standard error.

Hence taking $\frac{1369}{3625}$ (or 37.77%) to be the estimate of the Sanskrit words in the writer's vocabulary, we have that the limits are 37.77 per cent ±3(.81) per cent, *i.e.*, 35.34 per cent and 40.20 per cent.

REVISIONAL PROBLEMS

PROBLEM 14.35. In 324 throws of a six-faced die odd points appeared 181 times. Would you say that the die is 'fair' ?

$$\left[\textit{Ans.}\ \frac{\text{Difference}}{\text{S. E.}}=\frac{19}{9}=2.1\right]$$

PROBLEM 14.36. 160 heads and 240 tails resulted from 400 tosses of a coin. Does this coin appear to be unbiased ?

$$\left[\textit{Ans.}\ \frac{\text{Difference}}{\text{S.E.}}=\frac{40}{10}=4\right]$$

PROBLEM 14.37. In a random sample of 1,000 persons from town A, 400 are found to be consumers of wheat. In a sample of town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as proportion of wheat consumers is concerned.

$$\left[\textit{Ans.}\ \frac{\text{Difference}}{\text{S.E.}}=\frac{.1}{.028}=3.6\right]$$

PROBLEM 14.38. In an investigation about canning of fruits, it was found that of the 270 rural families enumerated 357 had done canning during the preceding season while among the 300 city families 274 had canned during the preceding season. Test the hypothesis that the proportion of families with previous canning experience is the same in urban and rural communities.

$$\left[\textit{Ans:}\ \frac{\text{Difference}}{\text{S.E.}}=\frac{.052}{.181}=2.9\right]$$

PROBLEM 14.39. There are 1000 students in a college. Out of 20,000 in the whole university, in a study 200 were found smokers in the college and 1000 in the whole University. Is there a significant difference between the proportions of smokers, in the college and university ?

$$\left[\textit{Ans:}\ \frac{\text{Difference}}{\text{S.E.}}=\frac{0.15}{.0068}=22.06\right]$$

PROBLEM 14.40. A coin is tossed 576 times and the number of heads observed is 280. Can the coin be said to be unbiased ?

$$\left[\textit{Ans:}\ \frac{\text{Difference}}{\text{S.E.}}=\frac{8}{12}=.67\right]$$

PROBLEM 14.41. In a sample of 500 villages in Rajasthan, 280 are found to be wheat eaters and the rest rice eaters. Can we assume that both the food articles are equally popular ?

$$\left[\textit{Ans:}\ \frac{\text{Difference}}{\text{S.E.}}=\frac{30}{11.2}=2.68\right]$$

PROBLEM 14.42. A railway company installed two sets of 50 Burma ties each. The two sets were treated with creosote by two different processes. After a number of years of service, it was found that 22 ties of the first set and 18 ties of the second set were still in good condition. Are we justified in claiming that there is no real difference between the preserving properties of the two processes ?

$$\left[\textit{Ans.}\ \frac{\text{Difference}}{\text{S.F.}} = \frac{.08}{.098} = .82 \right]$$

PROBLEM 14.43. Suppose four coins are tossed simultaneously 1,600 times and falling of heads are called successes. We have thus 1,600 samples of four tosses each. The following distribution has been obtained :—

Success	Frequency
4	90
3	420
2	580
1	410
0	100
Total	1,600

$$\left[\textit{Ans.}\ \frac{\text{Difference}}{\text{S.E.}} = \frac{.006}{1} = .006 \right]$$

PROBLEM 14.44. A group of scientific men reported 1,705 sons and 1,527 daughters. Do these figures conform to the hypothesis that the sex ratio is $\frac{1}{2}$?

$$\left[\textit{Ans.}\ \frac{\text{Difference}}{\text{S.E.}} = \frac{0.0275}{.0088} = 3.13 \right]$$

Sampling of Variables—Large Samples

15

In this chapter questions have been solved on sampling of variables. Only large samples have been considered : most of the questions solved relate to the calculation of the standard error of various statistical measures. Here also it has been assumed that the distributions are fairly normal and practically all the sample means would fall within Mean ± 3 standard error. The standard error of various measures have been calculated by using the following formulae:—

(1) Standard Error of Mean $= \frac{\sigma p}{\sqrt{N}}$ or $\frac{\sigma \text{ (sample)}}{\sqrt{N}}$

(2) ,, ,, ,, Median $= 1.25331 \frac{\sigma}{\sqrt{N}}$

(3) ,, ,, ,, Quartiles $= 1.36263 \frac{\sigma}{\sqrt{N}}$

(4) ,, ,, ,, Mean Deviation $= .6028 \frac{\sigma}{\sqrt{N}}$

(5) ,, ,, ,, Semi Inter Quartiles Range $= .78672 \frac{\sigma}{\sqrt{N}}$

(6) ,, ,, ,, Standard Deviation $= \frac{\sigma}{\sqrt{2N}}$

(7) ,, ,, ,, Co-efficient of Skewness $= \sqrt{\frac{3}{2N}}$

(8) ,, ,, ,, Variance $= \sigma^2 \sqrt{\frac{2}{N}}$

(9) ,, ,, ,, Coefficient of Variation $= \frac{V}{\sqrt{2N}} \sqrt{1 + \frac{2V^2}{10^4}}$

(10) ,, ,, ,, Co-efficient of Correlation $= \frac{1-r^2}{\sqrt{N}}$

(11) ,, ,, ,, Regression Co-efficient $= \frac{\sigma_x \sqrt{1-r^2}}{\sigma_y \sqrt{N}}$

(12) ,, ,, ,, Regression estimate $= \frac{\sigma_x \sqrt{1-r^2}}{\sigma_y \sqrt{1-r^2}}$

(13) ,, ,, ,, Co-efficient of Association=

$$\frac{1-Q^2}{2} \sqrt{\frac{1}{(AB)} + \frac{1}{(A\beta)} + \frac{1}{(\alpha B)} + \frac{1}{(\alpha\beta)}}$$

Most of the problems in this chapter relate to testing the significance of a sample mean, median or standard deviation, etc., and finding out whether it significantly differs from population mean, median or standard deviation, etc. Problems have also been solved on setting limits to different statistical measures.

STANDARD ERROR (OF THE MEAN)

PROBLEM 15.1. *Calculate the standard error of the mean from the following data collected in one of the many random sample inquiries conducted to find out average earnings of a particular class :—*

Earning Per week	*Number of Persons*
up to 10	*50*
,, ,, *20*	*150*
,, ,, *30*	*300*
,, ,, *40*	*500*
,, ,, *50*	*700*
,, ,, *60*	*800*
,, ,, *70*	*900*
,, ,, *80*	*1000*

SOLUTION. Calculation of the standard deviation of the earnings.

Earning per week in rupees (*m*)	Mid-values (*mv*)	Number of persons (*f*)	Deviation from the as. av. (45) (*dx*)	*dx*/*i*, where *i* = 10	Frequency × deviation *fdx*	Square of the deviation *dx*²	Frequency × Square of the deviation *fdx*²
00—10	5	50	—40	—4	—200	16	800
10—20	15	100	—30	—3	—300	9	900
20—30	25	150	—20	—2	—300	4	600
30—40	35	200	—10	—1	—200	1	200
40—50	45	200	0	0	0	0	0
50—60	55	100	+10	+1	+100	1	100
60—70	65	100	+20	+2	+200	4	400
70—80	75	100	+30	+3	+300	9	900
		n = 1,000			Σfdx = —400		$\Sigma fdx^2 = 3{,}900$

Substituting the above values in the formula

$$\sigma = \sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)} \times i$$

Where σ represents the standard deviation, we get,

$$\sigma = \sqrt{\frac{3900}{1000} - \left(\frac{-400}{1000}\right)^2} \times 10 = 1.934 \times 10 = \text{Rs. } 19.34$$

Substituting the values of σ and N in the formula

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \;;$$

Where σ_m represents the standard error of the mean, we get,

$$\sigma_m = \frac{19.34}{\sqrt{1000}} = \cdot 61$$

PROBLEM 15.2. *To know the mean weight of all 10-year-old boys in the state of Rajasthan—a sample of 225 is taken. The mean weight of this sample is found to be 67 pounds with a standard deviation of 12 pounds. Can you draw any inference from it about the mean weight of the universe ?*

SOLUTION. Standard error of the sample mean (weight) is

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{12}{\sqrt{225}} = .8 \text{ pound.}$$

Assuming that the conditions of simple sampling hold good the mean weight of the universe in all probability would be lying within the limits set by three times this standard error to the mean of the sample.

So, the mean weight of all 10-year-old boys in the State of Rajasthan lies between 67 pounds +2.4 pound, that is 64.6 pounds and 69.4 pounds.

PROBLEM 15.3. *The following frequency table shows the distribution of 1110 observations made on 149 commodity price series during ten business cycles.*

Duration af cycle (in months)	*Frequency*
7.50 to 12.49	*7*
12.50 to 17.49	*27*
17.50 to 22.49	*61*
22.50 to 27.49	*115*
27.50 to 32.49	*139*
32.50 to 37.49	*186*
37.50 to 42.49	*167*
42.50 to 47.49	*124*
47.50 to 52.49	*122*
52.50 to 57.49	*67*
57.50 to 62.49	*52*
62.50 to 76.49	*15*
67.50 to 72.49	*15*
72.50 to 77.49	*8*
77.50 to 82.49	*2*
82.50 to 87.49	*2*
87.50 to 92.49	*0*
92.50 to 97.49	*1*

Calculate the average duration of the cycle for this distribution and state that within what limits is your average correct ?

SOLUTION. Computation of the average duration of the cycle and its standard deviation.

Duration of cycle in months	Mid-points	Fre-quency	Deviation from the as. av. (40) taking $i=5$ as common		Square of deviation	Frequency × deviation	Frequency × Square of deviation
(m)	(mp)	(f)	(dx)	(dx/i)	dx^2	fdx	fdx^2
7.50—12 ·49	10	7	—30	— 6	36	— 42	252
12.50—17. 49	15	27	—25	— 5	25	—135	675
17.50—22.49	20	61	—20	— 4	16	—244	976
22.50—27.49	25	115	—15	— 3	9	—345	1035
27.50—32 49	30	139	—10	— 2	4	—278	556
32.50—37·49	35	186	— 5	— 1	1	—186	186
37.50—42·49	40	167	0	0	0	0	0
42.50—47·49	45	124	+ 5	— 1	1	+124	124
47.50—52·49	50	122	+10	+ 2	4	+244	488
52.50—57·49	55	67	+15	+ 3	9	+201	603
57.50—62·49	60	52	+20	+ 4	6	+208	832
62.50—67·49	65	15	+25	+ 5	25	+ 75	375
67·50—72.49	70	15	+30	+ 6	36	+ 90	540
72.50—77.49	75	8	+35	+ 7	49	+ 56	392
77.50—82.49	80	2	+40	+ 8	64	+ 16	128
82.50—87.49	85	2	+45	+ 9	81	+ 18	162
87.50—92.49	90	0	+50	+10	100	0	0
92.50—97.49	95	1	+55	+11	121	+ 11	121
		$n=$ 1110				$\Sigma fdx=$ —187	$\Sigma fdx^2=$ 7445

Mean duration of cycle of $a = x + \frac{\Sigma fdx}{n} \times i = 40 + \frac{-187}{1110} \times 5$

$= 39.16$ months

Standard deviation of the duration of cycle is

$$\sigma = \sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i = \sqrt{\frac{7445}{1110} - \left(\frac{-187}{1110}\right)^2} \times 5$$

$= 2.58 \times 5 = 12.9$ months.

Substituting the values of σ and N in the formula

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

Where σ_m represents the standard error of the mean

We get,

$$\sigma_m = \frac{12.9}{\sqrt{1110}} = .39 \text{ months.}$$

The average duration of the cycle for the given distribution is 39.16 months. But there is every possibility that this sample study may be loaded with cycles of shorter duration or of longer duration. So this sample mean cannot be the true average of the universe. However, in all probability the calculated average is correct within the limits set to it by three times its standard error (that is, 1.17 months) or 39.16 +1.17 months.

PROBLEM 15.4. *An industry desires to make a survey of the mean weekly wage of 10,000 of its workers. Since a study of all the workers is impossible, a* *representative sample of 400 workers is selected. The mean weekly wage of the 400 workers is Rs. 30 and the standard deviation Rs. 2.50. If additional sample were selected, by how much would the results differ from the above sample ?*

SOLUTION. Standard error of the mean weekly wage of the given sample is

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{2.5}{\sqrt{400}} = .125$$

If additional samples were selected, their mean would not differ by the mean weekly wage of this sample by more than three standard error of the mean of this sample. That is, the mean weekly wages of all the additional samples would lie between Rs. $30 \pm .375$ or Rs. 29.625 and Rs. 30.375.

PROBLEM 15.5. *To test if a small electric current affected the growth of wheat seedling, 100 pairs of plants were grown in parallel boxes and one member of each pair was "treated" by receiving a small electric current. The mean difference in heights between the treated and the untreated (treated—untreated) is +4 m.m, with a standard deviation of 2 m.m.*

Does the electric current exercise any influence on the growth of the plants ?

SOLUTION. Let us presume that electric current has no effect on the growth of the plants and the mean of the universe is therefore, 0.

The standard error of the mean

$$= \frac{\sigma}{\sqrt{N}} = \frac{2}{\sqrt{100}} = .2 \text{ m.m.}$$

The actual difference of 4 *m.m.* is 20 times the standard error and as such significant. The hypothesis has to be rejected and we have to conclude that electrification exercises an influence on the growth of the plants.

PROBLEM 15.6. *Suppose that it has been determined that the average pulse rate of males in the 20-25 year age group is 72 beats per minute and that the standard deviation is 9.5 beats per minute. If a group of 55 distance runners, all in the given age-group, were examined and found to have an average pulse rate of 65, should this be regarded as a significant deviation from the general average ?*

SOLUTION. Standard error of the average pulse rate of 55 distance runners is

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{9.5}{\sqrt{55}} = 1.28 \text{ beats per minute}$$

The difference between the two average pulse rate is 7 beats per minute, which is 5.47 times this standard error.

Hence the deviation of the average pulse rate of distance runners from the general average is significant.

PROBLEM 15.7. *Suppose that for tyres of a certain factory, the mean kilometrage. is 15180 and the standard deviation is 1248 kilometres.*

Suppose further, that a sample of 900 tyres shows a mean kilometrage of 15232 kilometres with a standard deviation of 1210 kilometres. Ascertain if the sample mean represents a significant divergence from the population mean.

SOLUTION. The standard error of the sample mean is

$$\sigma_m = \frac{\sigma p}{\sqrt{N}}$$

where σp is the standard error of the entire population and N is the number of items in the sample

Substituting N in the number of items in the sample

Substituting the given values, we get

$$\sigma_m = \frac{1248}{\sqrt{900}} = 41.6 \text{ kilometres.}$$

The observed difference between the two means of kilometrage is 15232–15180, or 52 kilometers. This is only 1.25 times this standard error and so could have arisen through the fluctuations of simple sampling. Hence it can be said that the divergence of the sample mean from the population mean is not significant.

Note:—When the standard deviation of the population is given, the standard deviation of the sample should not be used as the latter is only a substitute for the forme..

PRCBI EM 15.8. *If the standard deviation of pulse rate in adults is 8 and the normal pulse rate is 70, would you say high pulse rate was diagnostic if a group of 64 people suffering from a disease were found to have a pulse rate of 75 ?*

SOLUTION. The standard error of the sample pulse rate is

$$\sigma_m = \frac{\sigma p}{\sqrt{N}}$$

Substituting the values, we get, $\sigma_m = \frac{8}{\sqrt{64}} = 1$

The deviation of the sample rate from the normal rate is 5 and it is 5 times the standard error—and as such highly significant. A high pulse rate was definitely diagnostic in the above case.

PROBLEM 15.9. *The following table shows frequency distribution of yield of wheat in quintals per acre in 998 irrigated fields selected at random in the province of Punjab.*

Limit in Qtls.	*No. of fields*
0— 4	*45*
4— 8	*184*
8—12	*281*
12—16	*228*
16—20	*155*
20—24	*77*
24—28	*22*
28—32	*5*
32—36	*1*

Calculate the sampling error of the mean.

SOLUTION. Calculation of the mean yield of wheat and standard deviation.

Yield of wheat per acre in Qtls. (*m*)	Mid-values (*m.v.*)	No. of fields (*f*)	Deviations from the as. av. (14) taking $i = 4$ as common (*dx*)	(*dx*/*i*)	Square of deviation dx^2	Frequency × deviation *fdx*	Frequency × square of deviation fdx^2
0— 4	2	45	—12	—3	9	—135	405
4— 8	6	184	— 8	—2	4	—368	736
8—12	10	281	— 4	—1	1	—281	281
12—16	14	228	0	0	0	0	0
16—20	18	155	+ 4	+1	1	+155	155
20—24	22	77	+ 8	+2	4	+154	308
24—28	26	22	+12	+3	9	+ 66	198
28—32	30	5	+16	+4	16	+ 20	80
32—36	34	1	+20	+5	25	+ 5	25
		$n = 998$				$\Sigma fdx =$ —384	$\Sigma fdx^2 =$ 2,188

Mean yield of wheat per acre $= x + \dfrac{\Sigma dx}{n} \times i$

$$= 14 + \frac{-384}{998} \times 4 = 12.46 \text{ Qtls.}$$

Standard deviation of the yield per acre is

$$\sigma = \sqrt{\frac{\Sigma f dx^2}{n} - \left(\frac{\Sigma f dx}{n}\right)^2} \times i = \sqrt{\frac{2188}{998} - \left(\frac{-384}{998}\right)^2} \times 4$$

$$= 1.43 \times 4 = 5.72 \text{ Qtls.}$$

Standard error of the mean yield of wheat per acre is

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{5.72}{\sqrt{998}} = .18 \text{ Qtls.}$$

The sampling error is equal to standard error multiplied by critical value. The critical value differs with the level of confidence. At 5% level of confidence the critical value is 1.96.

Hence, at this level of confidence, the sampling error of the mean yield of wheat per acre is 1.96 × .18 or .35 approx. However, if sampling error of the mean is required at 1% level of confidence, it would be 2.58 × .18 or .465 approx.

PROBLEM 15.10. *Suppose that the standard deviation of stature in men is 2.48 cms. One hundred male students in a large university are measured and their average height is found to be 168 .52 cms. Determine the 98 per cent confidence limits for the mean height of the men of the university.*

SOLUTION. Assuming that the conditions of simple sampling hold true, the standard error of the average height of 100 male students is

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{2.48}{\sqrt{100}} = .248 \text{ cms.}$$

Mean of the sample ± 2.32 standard error of the man gives the 98 per cent confidence limits for the mean of the universe. So, the 98 per cent confidence limits for the mean height of men of the university is 168.52 cms. ± .575 cms., that is 167.945 cms. and 169.095 cms. 98% confidence limits mean the same thing as 2% level of confidence the critical value of which is equal to 2.32.

PROBLEM 15.11. *The average breaking strength of steel rods is specified to be 18.5 thousand pounds. To test this a sample of 100 rods was tested and gave the following results ;—*

Breaking Strength	*No. of rods*
15,000—16,000	*12*
16,000—17,000	*20*
17,000—18,000	*40*
18,000—19,000	*20*
19,000—20,000	*8*

Do the results of the tests justify the hypothesis ?

SOLUTION. Calculation of the mean and standard deviation.

Breaking Strength (000) m	M. P.	No. of Rods (f)	Deviation from as. av. (17.5) (dx)	fdx	fdx^2
15—16	15.5	12	−2	−24	48
16—17	16.5	20	−1	−20	20
17—18	17.5	40	0	0	0
18—19	18.5	20	+1	+20	20
19—20	19.5	8	+2	+16	32
	N = 100			$\Sigma fdx = -8$	$\Sigma fdx^2 = 120$

$$\text{Mean}=17.5+\left(\frac{-8}{100}\right)=17.42 \text{ thousand pounds.}$$

$$\text{Standard Deviation}=\sqrt{\frac{120}{100}-\left(\frac{-8}{100}\right)^2}$$

$$=\sqrt{1.1936}=1.09 \text{ thousand pounds.}$$

$$\text{The standard error of the mean}=\frac{\sigma}{\sqrt{N}}=\frac{1.09}{\sqrt{100}}=.109$$

The actual difference of means $(18.5—17.42)=1.08$ thousand pounds, and this is more than thrice the standard error, hence, the difference is significant and hypothesis has to be rejected.

PROBLEM 15.12. *Explain what you understand by 'Standard error' of the mean of random sampling. The data concerning height measurement for a random sample of individuals from a given population are as follows :*

Mean is equal to	*172*
S..D ,, ,, ,,	*12*
N. ,, ,, ,, ,,	*65*

If a large number of sample of the same size were selected at random from the given population, what would be the limits of the 2% confidence interval for the true mean.

SOLUTION. Standard error of the mean height of the sample is

$$\sigma m=\frac{\sigma}{\sqrt{N}}=\frac{12}{\sqrt{65}}=1.5 \text{ approx.}$$

At 2% level of confidence, the limits of the variation of the mean would be $172\pm(1.5\times2.32)=172\pm3.48=168.52$ and 175.48.

PROBLEM 15.13. *A sample of 900 members is found to have a mean of 3.4 cm. Can it be reasonably regarded as a simple sample from a large population with mean 3.2 cm. and standard deviation 2.3 cm.*

SOLUTION. Assuming that the conditions of simple sampling hold good the standard error of the mean of the sample is

$$\sigma_m=\frac{\sigma}{\sqrt{N}}=\frac{2.3}{\sqrt{900}}=.077 \text{ cm.}$$

The difference between the sample mean and the population mean is .2 cm. which is only 2.6 times this standard error and so may have arisen through the fluctuations of simple sampling. So, it leads us to the conclusion, though not strongly that the given sample is a simple sample from the given population.

PROBLEM 15.14. *A certain colliery is supposed to supply coal of ash content about 15%. One hundred samples of coal of the company when analysed gave an ash content of 16.8% with a standard deviation of 8%. Do these results justify company's claim ?*

SOLUTION. The standard error of the percentage of ash contents in the above problem is

$$\sigma_m=\frac{\sigma}{\sqrt{N}}=\frac{8}{\sqrt{100}}=.8$$

The actual difference between the two percentages is 1.8 which is 2.25 times the standard error. Though technically such a difference could arise due to sampling fluctuations yet its chances are very very remote. The company's claim is not justified. (The theoretical maximum for sampling fluctuation is about three times the standard error but in actual practice the results are judged at twice the standard error or even less—say 1.96 times the standard error, which gives 5% confidence limits).

PROBLEM 15.15. *What is the standard error of the standard deviation if the parent distribution is (a) normal (b) not abnormal ?*

It is known that the mean and standard deviation of a variable are respectively 100 and 10 in the universe. It is, however, considered sufficient to draw a sample of sufficient size but such as to ensure that the mean of the sample would be, in all probability within 0.01% of the true value. How much would the cost be (exclusive of overhead charges) if the charges for drawing 100 members of a sample be one rupee ?

SOLUTION. Assuming that the conditions of the simple sampling hold, the mean of the sample cannot deviate from the mean of the universe by more than 3 times the standard error of the mean of the sample.

So, in a universe with a mean of 100, if the mean of the sample is to be within .01 limits of it, N should be such that

$$3\,\frac{\sigma}{\sqrt{N}}=.01 \quad \text{or} \quad 3\,\frac{10}{\sqrt{N}}=.01$$

Simplifying it, we get $\dfrac{30}{\sqrt{N}}=.01$

or $.01\sqrt{N}=30$; or $\sqrt{N}=3000$ or $N=90{,}00{,}000$

Thus, 90,00,000 members have to be drawn from the universe to get the given precision and at the given rate of charges, the total cost would be Rs. 90,000.

STANDARD ERROR (OF MEDIAN, M. D., S. D., ETC.)

PROBLEM 15.16. *Suppose it has been found out that the median weight of the soldiers of Rajput Regiment is 154.7 lbs. If the number of the soldiers is 7750 and the standard deviation of the weight is 21.3 lbs. find out the standard error of the median weight.*

SOLUTION. Standard error of the median weight.

$$=1.25331\,\frac{\sigma}{\sqrt{N}}$$

Substituting the values for σ and N.

We get,

$$s.e._{mdn}=\frac{1.25331\times 21.3}{\sqrt{7750}}=.303 \text{ lbs.}$$

Thus, the standard error of the median weight is .303 lbs.

PROBLEM 15.17. *Suppose it has been found out that the first quartile height of 580 past-gruduate students of a university is 165.71 cms. and the third quartile height for the same is 169.21 cms. with a standard deviation of 14.56 cms. Find out the standard errors of the two quartiles.*

SOLUTION. Standard errors of the first and the third quartiles are given by the formula,

$$s.e._q=1.36263\,\frac{\sigma}{\sqrt{N}}$$

Substituting the value of a and N, we get,

$$s.e._q=\frac{1.36263\times 14.56}{\sqrt{580}}$$

$=$Anti log. $[(\log. 1.36263+\log. 14.56-\tfrac{1}{2}(\log. 580)]$

$=$Anti log. $[(.1341+1.1632)-\tfrac{1}{2}(2.7634)]$

$=$Anti log. $[(1.2973-1.3817)]=$Anti log. $\bar{1}.9156=.8233$

Thus, the standard error of the quartiles is .8233.

PROBLEM 15.18. *Suppose in U. S. A., 250000 marriages were contracted in a certain year in which the ages of the bridegrooms varied from 20 years to 80 years. If this distribution has got a standard deviation of 8 years, what would be the standard error of the Ist, and 3rd, etc., decile ages of the bridegrooms if they are calculated from the given data.*

SOLUTION. The standard errors cf the 1st and the 9th decile are given by the formula.

$$s.e.\ \text{1st or 9th Decile} = 1.70942 \frac{\sigma}{\sqrt{N}}$$

Substituting the values of σ and N, we get for the age of bridegrooms.

$$s.e.\ \text{1st or 9th Decile} = \frac{1.70942 \times 8}{\sqrt{250000}} = .02735 \text{ or } .027 \text{ year.}$$

The standard errors of the 2nd and the 8th decile is given by the formula

$$s.e.\ \text{2nd or 8th Decile} = 1.42877 \frac{\sigma}{\sqrt{N}}$$

Substituting the values of σ and N we get for the age of bridegrooms

$$s.e.\ \text{2nd or 8th Decile} = \frac{1.42877 \times 8}{\sqrt{250{,}003}} = .023 \text{ year.}$$

The standard error of the 3rd and the 7th deciles is given by the formula

$$s.e.\ \text{3rd or 7th Decile} = 1.31800 \frac{\sigma}{\sqrt{N}}$$

Substituting the values of σ and N we get for the age of the bridegrooms

$$s.e.\ \text{3rd or 7th Decile} = \frac{1.31800 \times 8}{\sqrt{250000}} = 0.21 \text{ year.}$$

The standard error of the 4th and the 6th deciles is given by the formula

$$s.e.\ \text{4th or 6th Decile} = 1.26804 \frac{\sigma}{\sqrt{N}}$$

Substituting the values of σ and N, we get for the age of bridegrooms

$$s.e.\ \text{4th or 6th Decile} = \frac{1.26804 \times 8}{\sqrt{250000}} = .020 \text{ year.}$$

The standard error of the 5th decile is the same as of the median and is given by the formula

$$s.e._{\text{mdn}} = 1.25331 \frac{\sigma}{\sqrt{N}} \quad ;$$

Substituting the values, we get,

$$s.e._{\text{mdn}} = \frac{1.25331 \times 8}{\sqrt{250000}} = .020 \text{ year.}$$

PROBLEM 15.19. *A sample of 100 is taken from the universe having a standard deviation of 7.6. The mean deviation of the sample is found to be 6.5. Calculate the standard error of this mean deviation.*

SOLUTION. The standard error of the mean deviation is given by the formula

$$s.\ e_{\text{md}} = .6028 \frac{\sigma}{\sqrt{N}}$$

Substituting the given values of σ and N we get,

$$s.e._{\text{md}} = \frac{.6028 \times 7.6}{\sqrt{100}} = .458.$$

Thus, the standard error of the mean deviation of the sample is .458.

PROBLEM 15.20. *The following table gives the frequency distribution of weights for adult males born in Punjab.*

Weight in lbs.	*Number of Men within given limits of weight*
90 –	*2*
100—	*34*
110—	*152*
120—	*390*
130—	*867*
140—	*1623*
150 –	*1559*
160—	*1326*
170—	*787*
180—	*476*
190—	*263*
200—	*107*
210—	*85*
220—	*41*
230—	*16*
240—	*11*
250—	*8*
260—	*1*
270—	*0*
280—	*1*
Total	*7749*

Find out the semi-interquartile range and calculate its standard error.

SOLUTION. Calculation of the semi-interquartile range and the standard deviation of the weights of adult males born in Punjab.

Weight in lbs.	Mid-values	Number of men	Cumulative Frequency	Deviation from the as. av. (165) taking i = 10 as common		Square of deviation	Frequency × deviation	Frequency × square of deviation
(m)	(mv)	(f)	$(c.f.)$	(dx)	(dx/i)	(dx^2)	(fdx)	(fdx)
90—	95	2	2	— 70	— 7	49	— 14	98
100—	105	34	36	— 60	— 6	36	— 204	1224
110—	115	152	188	— 50	— 5	25	— 760	3800
120—	125	390	578	— 40	— 4	16	—1560	6240
130—	135	867	1445	— 30	— 3	9	—2601	7803
140—	145	1623	3068	— 20	— 2	4	—3246	6492
150—	155	1559	4627	— 10	— 1	1	—1559	1559
160—	165	1326	5953	0	0	0	0	0
170—	175	787	6740	+ 10	+ 1	1	+ 787	787
180—	185	476	7216	+ 20	+ 2	4	+ 952	1904
190—	195	263	7479	+ 30	+ 3	9	+ 789	2367
200—	205	107	7586	+ 40	+ 4	16	+ 428	1712
210—	215	85	7671	+ 50	+ 5	25	+ 425	2125
220—	225	41	7712	+ 60	+ 6	36	+ 246	1476
230—	235	16	7728	+ 70	+ 7	49	+ 112	784
240—	245	11	7739	+ 80	+ 8	64	+ 88	704
250—	255	8	7747	+ 90	+ 9	81	+ 72	648
260—	265	1	7748	+100	+10	100	+ 10	100
270—	275	0	7748	+110	+11	121	+ 0	0
280—	285	1	7749	+120	+12	144	+ 12	144
		$n = 7749$					$\Sigma fdx =$ —6023	$\Sigma fdx^2 =$ 39967

First Quartile Weight

Q_1= the value of $\frac{n}{4}$ that is, $\frac{7749}{4}$ or 1937·25th item, which lies in the 140—150 weight group; By interpolation,

$$Q_1 = l_1 \frac{l_2-l_1}{f_1}(q_1-c)$$

$$= 140 + \frac{150-140}{1623}(1937{\cdot}25-1445) = 143.03 \text{ lbs.}$$

Third Quartile Weight.

Q_2= the value of $\frac{3(n)}{4}$ that is 5811.75th item.

which lies in 160—170 weight group

By interpolation,

$$Q_3 - 160 + \frac{170-160}{1326}(5811{\cdot}75-4627) = 168.63 \text{ lbs.}$$

Substituting the values of Q_3 and Q_1 in the formula

Semi-interquartile range $= \frac{Q_3-Q_1}{2}$; We get

Semi-interquartile range $= \frac{168.63-143.03}{2} = 12.80$ lbs.

Standard deviation of the weight is

$$\sigma = \sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} \times i$$

Substituting the values we get,

$$\sigma = \sqrt{\frac{39967}{7749} - \left(\frac{-6023}{7749}\right)^2} \times 10 = 2.133 \times 10 = 21.33 \text{ lbs.}$$

The standard error of the semi-interquartile range is given by the formula

s.e. Semi-interquartile range $.78672 \ \frac{\sigma}{\sqrt{N}}$

Substituting the values of σ and N, we get,

s.e. Semi-interquartile range $= \frac{.78672 \times 21.33}{\sqrt{7749}} = .187$ lbs.

Thus, the semi-interquartile range of the weights of the adult males born in Punjab is 12.96 lbs, and its standard error is .187 lb.

PROBLEM 15.21. *8500 males of the 20—25 age group were examined and were found to have a pulse rate of 71 beats per minute. The standard deviation is 9.5 beats per minute. Find out the standard error of this standard deviation assuming the universe to be normal.*

SOLUTION. If the universe is normal, the standard error of the standard deviation is given by the formula

$$s.e.\ \sigma = \frac{\sigma}{\sqrt{2N}}$$

Substituting the values of σ and N, we get

$$s.e.\ \sigma = \frac{9.5}{\sqrt{2 \times 8500}} = .073 \text{ approx.}$$

Thus, the standard error of the standard deviation of the pulse beat rate is .073 beat per minute.

PROBLEM 15.22. *For the height distribution of 7575 females, the standard deviation is found to be 6.46. cms. Within what limits this standard deviation may be taken to be correct ?*

SOLUTION. Taking the universe to be normal, the standard error of standard deviation of the height is

$$s.e.\ \sigma = \frac{\sigma}{\sqrt{2N}} = \frac{6.46}{\sqrt{2 \times 7575}} = 0.05 \text{ approx.}$$

Hence, assuming that the sampling is simple, we may say the *s. d.* in the universe almost certainly lies in the range 6.46±0.15 (*i. e.* Standard Deviation ±3 *s. e.* σ)

PROBLEM 15.23. *To find out the average age of bridegrooms in U. S. A. 202,857 marriage cases were studied. The average age of bridegrooms for this sample was found to have a standard deviation of 2.6556. Estimate the standard deviation for the average age in the population and assign limits within which this will probably lie ?*

SOLUTION. Assuming the parent distribution for the age of bridegrooms to be normal, the standard error of the standard deviation of the sample is

$$s.e.\ \sigma = \frac{\sigma}{\sqrt{2N}} = \frac{2.6556}{\sqrt{202857 \times 2}} = .0042 \text{ approx.}$$

If the sampling is simple, the standard deviation of the universe cannot deviate from the standard deviation of the sample by more than three times the standard error of the standard deviation of the sample. Hence, we may say that the *S. D.* in the universe almost certainly lies in the range 2.6556±0.0126, that is 2.6430 and 2.6682.

PROBLEM 15.24. *The Pearson's co-efficient of skewness* $\left(\frac{a-z}{\sigma}\right)$ *for a distribution is +.12. Find out its standard error if the number of items in the distribution are 150.*

SOLUTION. The standard error of the co-efficient of skewness

$$s.e.j = \sqrt{\frac{3}{2N}}$$

Substituting the values we get,

$$s.e.j = \sqrt{\frac{3}{2 \times 150}} = \sqrt{\frac{3}{300}} = \sqrt{.01} = .1.$$

Thus, the standard error of the Co-efficient of skewness is .1.

PROBLEM 15.25. *Calculate the co-efficient of skewness of the following distribution and also calculate its standard error.*

Size of item	*Frequency*
3	*2*
4	*4*
5	*5*
6	*8*
7	*10*
8	*7*
9	*8*
10	*6*

SOLUTION. Calculation of arithmetic average, mode and standard deviation.

Size of item (*m*)	Frequency (1) (*f*)	Deviation from as. av. (7) (*dx*)	Square of deviation (dx^2)	Frequency × deviation (*fdx*)	Frequency × Square of deviation (fdx^2)	Frequency (2)	(3)	(4)	(5)	(6)
3	2	−4	16	−8	32					
4	4	−3	9	−12	36	6				
5	5	−2	4	−10	20		9	11		
6	8	−1	1	−8	8	13			17	
7	10	0	0	0	0		18			23
8	7	+1	1	+7	7	17		25		
9	8	+2	4	+16	32		15		25	
10	6	+3	9	+18	54	14				21
	N=50			Σfdx =+3	Σfdx^2 =189					

Arithmetic average, or $a = x + \frac{\Sigma dx}{n} = 7 + \frac{3}{50} = 7.06.$

Standard deviation.

$$\sigma = \sqrt{\frac{\Sigma fdx^2}{n} - \left(\frac{\Sigma fdx}{n}\right)^2} = \sqrt{\frac{189}{50} - \left(\frac{3}{50}\right)^2} = 1.94$$

Analysis Table for the Calculation of Mode.

Column	Size of item containing maximum frequency				
(1)			7		
(2)			7	8	
(3)		6	7		
(4)		6	7	8	
(5)			7	8	9
(6)	5	6	7		
No. of times the size of item occurs	1	3	6	3	1

From the above table, it is clear that 7 is the size of item which is most frequented. It is not true of any other size. Hence, the mode, or *z* is 7.

Substituting these values in the formula

$$j = \frac{a - z}{\sigma}$$

where *j* represents the co-efficient of skewness,

We get

$$j=\frac{7.06-7}{1.94}=.03.$$

The standard error of the co-efficient of skewness is

$$s.\,e.\,j=\sqrt{\frac{3}{2N}}$$

Substituting the values, we get

$$s.e.j=\sqrt{\frac{3}{2\times 50}}=.1732.$$

Thus, the co-efficient of skewness is .03.

and its standard error is ·1732.

Note.—This formula of measuring standard error of the co-efficient of skewness is applicable only when the co-efficient of skewness is measured by $\frac{a-z}{\sigma}$ as above.

STANDARD ERROR (OF C. V., 'r' Reg. Etc.)

PROBLEM 15.26. *The following is the distribution of Bradstreet's commodity prices (1957-1973), expressed as percentages of trend.*

Percentage of Trend	*Frequency*	*Percentage of Trend*	*Frequency*
85%	1	100%	14
86	0	101	30
87	2	102	15
88	0	103	15
89	4	104	12
90	2	105	12
91	1	106	4
92	3	107	7
93	6	108	2
94	10	109	0
95	8	110	3
96	13	111	1
97	5	112	1
98	16	113	1
99	15	114	1

(a) What would be the limits of error in the calculation of the above distribution.

(b) What is the probable error for the Second Moment about the mean for the data of the distribution ?

SOLUTION. (*a*) Calculation of the variance and the standard deviation.

Percentage of Trend (m)	Frequency (f)	Deviations from the as. av. (100) (dx)	Square of deviation (dx^2)	Frequency × deviation (fdx)	Frequency × Square of deviation (fdx^2)
85	1	—15	225	—15	225
86	0	—14	196	0	0
87	2	—13	169	—26	338
88	0	—12	144	0	0
89	4	—11	121	—44	484
90	2	—10	100	—20	200
91	1	— 9	81	— 9	81
92	3	— 8	64	—24	192
93	6	— 7	49	—42	294
94	10	— 6	36	—60	360
95	8	— 5	25	—40	200
96	13	— 4	16	—52	208
97	5	— 3	9	—15	45
98	16	— 2	4	—32	64
99	15	— 1	1	—15	15
100	14	0	0	0	0
101	30	+ 1	1	+30	30
102	15	+ 2	4	+30	60
103	15	+ 3	9	+45	135
104	12	+ 4	16	+48	192
105	12	+ 5	25	+60	300
106	4	+ 6	36	+24	144
107	7	+ 7	49	+49	343
108	2	+ 8	64	+16	128
109	0	+ 9	81	0	0
110	3	+10	100	+30	300
111	1	+11	121	+11	121
112	1	+12	144	+12	144
113	1	+13	169	+13	169
114	1	+14	196	+14	196
	n=204			Σfdx= —12	Σfdx= 4968

Substituting the above values in the formula

$$\text{Variance}=\sigma^2=\frac{fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2$$

We get,

$$\text{Variance}=\frac{4968}{204}-\left(\frac{-12}{204}\right)^2$$

$$=24.35-.003=24.347.$$

The standard error of the variance is

$$s.e._{V}=\sigma^2\sqrt{\frac{2}{n}}=24.347\sqrt{\frac{2}{204}}=2.34$$

Thus, the limits of the variance are $24.347 \pm (3\times 2.34)$ or 24.347 ± 7.02.

(*b*) From the above, the standard error of the variance is 2.24.

Since variance=2nd Moment.

$\therefore$ the Probable error of the second moment $=.6745\times 2.34.=1.58.$

PROBLEM 15.27. *What would be the standard error of the variance, if, in ries the number of items are 100 and the standard deviation is 5.*

SOLUTION, Variance$=\sigma^2$

In this problem variance$=(5)^2=25$

The standard error of variance

$$=\sigma^2\sqrt{\frac{2}{n}}$$

Substituting the values, we get,

$$s.e._V=25\sqrt{\frac{2}{100}}=3.535$$

Thus the standard error of the variance in the problem is 3.535.

PROBLEM 15.28. *The following is the frequency table of six Months Prime Commercial Paper Rates, January 1962 to December, 1971.*

Class Intervals	*Frequencies*
2.50 to 2.99%	*2*
3.00 to 3.49 ,,	*7*
3.50 to 3.99 ,,	*20*
4.00 to 4.49 ,,	*30*
4.50 to 4.99 ,,	*20*
5.00 to 5.49 ,,	*10*
5.50 to 5.99 ,,	*5*
6.00 to 6.49 ,,	*6*

Calculate the standard error of the co-efficient of variation for the above able.

SOLUTION. Calculation of the co-efficient of variation.

Size of item	Mid-values	Frequency	Deviation from the as. av. (4.25)	Square of deviation	Frequency × deviation	Frequency × square of deviation
2.50 to 2.99%	2.75%	2	—1.50	2.25	—3	4.50
3.00 to 3.49 ,,	3.25 ,,	7	—1.00	1.00	—7	7.00
3.50 to 3.99 ,,	3.75 ,,	20	—0.50	0.25	—10	5.00
4.00 to 4.49 ,,	4.25 ,,	30	0	0	0	0.00
4.50 to 4.99 ,,	4.75 ,,	20	+0.50	0.25	+10	5.00
5.00 to 5.49 ,,	5.25 ,,	10	+1.00	1.00	+10	10.00
5.50 to 5.99 ,,	5.75 ,,	5	+1.50	2.25	+7.5	11.25
6.00 to 6.49 ,,	6.25 ,,	6	+2.00	4.00	+12	24.00
	$n=100$				$\Sigma fdx=+19.5$	$\Sigma fdx^2=66.75$

Arithmetic average of the distribution is

$$\sigma=x+3\frac{\Sigma fdx}{n}=4.25+\frac{19.5}{100}=4.445 \text{ (approx).}$$

Standard deviation is

$$\sigma=\sqrt{\frac{\Sigma fdx^2}{n}-\left(\frac{\Sigma fdx}{n}\right)^2}$$

$$=\sqrt{\frac{66.75}{100}-\left(\frac{19.5}{100}\right)^2}=.79$$

Co-efficient of variation is

$$V=\frac{\sigma\times100}{\sigma}=\frac{.79\times100}{4.4}=17.9$$

The standard error of the co-efficient of variation is given by the formula

$$s.e.\ \sigma_V=\frac{V}{\sqrt{2n}}\sqrt{1+\frac{2V^2}{10^4}}$$

Substituting the values of V and N in the formula

$$s.e.\ \sigma_V=\frac{17.95}{\sqrt{2\times100}}\sqrt{1+\frac{2\times(17.95)^2}{10^4}}=1.3$$

Thus, the standard error for the co-efficient of variation for the distributio is 1.3.

PROBLEM 15.29. *To study the correlation between the stature of the fath and the stature of the son, a sample of 900 is taken from the universe of fathers a sons. The sample study gives the correlation between the two to be + .67. With what limits does it hold true for the universe ?*

SOLUTION. The standard error of the correlation co-efficient between th stature of the father and son is

$$s.e._r=\frac{1-r^2}{\sqrt{N}}=\frac{1-(0.67)^2}{\sqrt{900}}=.018$$

If the sampling was simple, tne correlation in the universe cannot devi from the correlation in the sample by more than thrice this standard error, i. .054. Hence, correlation in the universe most probably lies within .67±.0 that is, .616 and .724.

PROBLEM 15.30. *To find out the correlation between the yield of milk p week and percentage of butter fat, a sample study of 12250 buffaloes is ma This gives the value of r to be—.082. Find out the standard error of the estimat*

SOLUTION. The standard error of the calculated r in the sample is

$$s.e._r=\frac{1-r^2}{\sqrt{N}}=\frac{1-(-0.082)^2}{\sqrt{12250}}=0.0089.$$

The correlation observed is nine times of its standard error and sm though it is, could not have arisen from sampling fluctuation.

PROBLEM 15.31. *Three experiments were conducted to determine relationship between laterality of hand and laterality of eye. The correlati between (1) difference of strength of grip and (2) difference in visual acuity were*

—0.02410 (3234 subjects)
—0.00738 (4003 subjects)
+0.02962 (1447 subjects)

Find the standard errors of the three correlation co-efficients, and hence sh that it cannot be concluded that there is any significant correlation between la rality of hand and laterality of eye.

SOLUTION. Thr standard error of the co-efficient of correlation is

$$s.e._r=\frac{1-r^2}{\sqrt{N}}$$

Where r represents the co-efficients of correlation and N, the total number of observations in an experiment.

Substituting the respective values of the three experiments, we get,

$s.e._r$ for the 1st experiment $= \dfrac{1-(0.02410)^2}{\sqrt{3234}} = 0.0176.$

$s.e._r$ for the 2nd experiment $= \dfrac{1-(0.00738)^2}{\sqrt{4003}} = 0.0158.$

$s.e._r$ for the 3rd experiment $= \dfrac{-(0.02962)}{\sqrt{1447}} = 0.0263.$

Since in all the three experiments of the co-efficient of correlation is less than thrice its standard error it is concluded that the correlation is not significant.

PROBLEM 15.32. *From the following four samples giving the stature in cms. for 400 adult males, 100 in England, 100 in Scotland, 100 in Wales and 100 in Ireland, find out as to what would be the standard error of the mean of a sample which consists of all these 400 men.*

	Stature in cms for adult males born in			
	England	*Scotland*	*Wales*	*Ireland*
Mean	*167.31*	*168.55*	*166.62*	*167.78*
Standard Deviation	*2.56*	*2.50*	*2.35*	*2.17*

SOLUTION. The combined mean of the samples or

$$m_{1234} = \frac{(167.31\times100)+(168.55\times100)+(166.62\times100)+(167.78\times100)}{100+100+100+100}$$

$$= 167.565$$

The combined standard deviation of 4 samples

$$\sigma_{1234} = \sqrt{\frac{n_1\sigma_1{}^2+n_1(m_1-m_{1234})^2+n_2\sigma_2{}^2+n_2(m_2-m_{1234})^2+n_3\sigma_3{}^2+n_3(m_3-m_{1234})^2+n_4\sigma_4{}^2+n_4(m_4-m_{1234})^2.}{n_1+n_2+n_3+n_4}}$$

$$= \sqrt{\frac{655.36+6.25+625.00+98.01+552.25+88.36+470.89+4.84}{400}}$$

$$= 2.50$$

Standard Error of the combined mean

$$= \frac{\sigma_{1234}}{\sqrt{N}} = \frac{2.5}{\sqrt{400}} = \frac{2.5}{20} = .125$$

PROBLEM 15.33. *Find out the regression co-efficient of Hapur prices over Allahabad prices from the following data and also calculate its standard error.*

Assuming n to be 900	*Hapur* *Rs.*	*Allahabad* *Rs.*
Av. price per quintal of wheat	*90*	*110*
Standard deviation	*3*	*2*

r between prices at Hapur and Allahabad+.67.

SOLUTION. Regression co-efficient of Hapur prices (x) over Allahabad Prices (y)

$$bxy = r\frac{\sigma x}{\sigma y} = .67\frac{3}{2} = 1.005$$

The standard error of this regression co-efficient is

$$s.e._{bxy}=\frac{\sigma x\sqrt{1-r^2}}{\sigma y\sqrt{N}}=\frac{3\sqrt{1-(.67)^2}}{2\sqrt{900}}=.037.$$

Thus, the regression co-efficient of Hapur prices over Allahabad prices is 1.005 and its standard error is .037.

PROBLEM 15.34. *To a certain number of students chosen at random from a given school population, a test of general intelligence (x) and a test of silent reading comprehension (y) are administered. Suppose that the following measures are derived from the distribution of scores on these two tests :—*

$Ax=102$; $\sigma x=12$; $Ay=80$; $\sigma y=10$
and $rxy=+.8$.

With the help of this data, estimate the scores that a student would make on general intelligence if he makes a score of 86 on silent reading comprehension test. Also calculate the standard error of your estimate.

SOLUTION. The regression equation of intelligence test (x) on reading test (y) is

$$(X-Ax)=r\,\frac{\sigma x}{\sigma y}(Y-Ay)$$

$$\text{or } X=r\frac{\sigma x}{\sigma y}(Y-Ay)+Ax$$

Substituting the given values, we get

$$X=\frac{.8\times 12}{10}(Y-80)+102$$

$$=.96Y-76.8+102$$

$$=.96Y+25.2$$

Estimating the scores on intelligence test (X) when scores on reading test (Y) is 86, we get

$X_{86}=.96\times 86+25.2=107.76$ scores.

The standard error of estimate of the scores on reading test is

$s.e._{xy}=\sigma x\sqrt{1-r^2}=12\sqrt{1-(.8)^3}=7.2$

Thus, if a student makes *a* score of 86 on reading test he would make a score of 108 on intelligence test. The standard error of the estimate is 7.2.

PROBLEM 15.35. *For a given group of adults, the co-efficient of correlation between height and weight is .6; standard deviation of height and of weight are 3 cms. and 12 lbs. respectively and the means in height and weight for the entire group are 169 cms. and 145 lbs, respectively. Find out the best estimate of the weight of an individual who is 172 cms. tall. Assign limits to this estimate in which in all probability his actual weight would be lying.*

SOLUTION. The regression of weight (Y) on height (X) is

$$(Y-Ay)=r\,\frac{\sigma y}{\sigma x}\,(X-A'x);\text{ or } Y=r\,\frac{\sigma y}{\sigma x}\,(X-A_1)+Ay.$$

Substituting the given values, we get

$$Y=\frac{.6\times 12}{3}\,(X-169)+145=2.4\,X-405.6+145=2.4\,X-260.6$$

Estimating the weight of an individual when his height is 172 cms, we get

$Y_{172}=2.4\times 172-260.6=152.2$ lb.

The standard error of this estimate is

$s.e._{yx}=\sigma y\sqrt{1-r^2}=12\sqrt{1-.6^2}=9.6$ lbs.

Thus, the best estimate of the weight of an individual who is 172 cms. tall is 152.2 lbs. and this estimate cannot deviate from his actual weight by more than three times this standard error. Hence, in all probability his actual weight would be lying between 152.2 lbs ±28.8 lbs., that is 123.4 lbs. and 181.0 lbs.

PROBLEM 15.36. *Two groups of children, one belonging to the professional class 125 in number, and the other belonging to the labouring class, [illegible]4 in number are compared and the following results are obtained :—*

	Poor children Per cent	*Well-to-do children Per cent*
Below normal weight	*55*	*13*
Above normal weight	*11*	*48*

Find the co-efficient of association between the weight of the children and their social status. Also calculate the standard error of this co-efficient of association.

SOLUTION. Denoting poor children by A, well-to-do children by a and below normal weight by B; and above normal weight by β.
We get,

$$(AB)=55 \quad (\alpha B)=13$$

$$(\alpha\beta)=48 \quad (A\beta)=11$$

Substituting these values in the formula,

$$Q=\frac{(AB)\,(\alpha\beta)-(A\beta)\,(\alpha B)}{(AB)\,(\alpha\beta)+(A\beta)\,(\alpha B)}$$

Where Q represents the co-efficient of association,
We get,

$$Q=\frac{(55\times48)-(11\times13)}{(55\times48)+(11\times13)}=\frac{2640-143}{2640+143}=+.89.$$

The standard error of this co-efficient of association is

$$s.e._{q}=\frac{1-Q^2}{2}\sqrt{\frac{1}{(AB)}+\frac{1}{(A\beta)}+\frac{1}{(aB)}+\frac{1}{(\alpha\beta)}}$$

$$=\frac{1-(.89)^2}{2}\sqrt{\frac{1}{55}+\frac{1}{11}+\frac{1}{13}+\frac{1}{48}}=.0472$$

Thus the co-efficient of association between the weight of the children and their social status is +.89 and its standard error is .0472.

REVISIONAL PROBLEMS-15

PROBLEM 15.37. A random sample of 1,000 farms in a certain year gives an average yield of wheat 2000 lb. per acre with a standard deviation of 192 lb. A random sample of 1,000 farms in the following year gives an average yield of 2,100 lb. per acre with a standard deviation of 202 lb. Are these data consistent with hypothesis that the average yields in the country were the same in these two years ?

$$\left[\textit{Ans}:\ \frac{\text{Difference}}{\text{S.E.}}=\frac{100}{8.8}=11.36\right]$$

PROBLEM 15.38. From a normal population, a random sample of size 32 was drawn and the sample standard deviation was found to be 1.38. Using 1 per cent level of significance, decide if it would be reasonable to adopt the value unity for the population standard deviation.

$$\left[\textit{Ans}:\ \frac{\text{Difference}}{\text{S. E.}}=\frac{1.38}{1.725}=.8\right]$$

PROBLEM 15.39. A correlation coefficient of 0.72 is obtained from a sample of 29 paired of observations. Is the difference significantly different from 0.8 ?

[*Ans* : 0.985]

PROBLEM 15.40. The standard deviation of the wage of 2,100 textile workers is Rs. 12.8: another sample of 600 workers gives standard deviation as Rs. 15.7. Find out if the standard deviation of the first sample significantly differs from the combined standard deviation of the two samples which is Rs. 14.0.

$$\left[Ans:\ \frac{2.9}{.46}=6.3\right]$$

PROBLEM 15.41. 490 male students and 450 female students appeared at the examination in statistics. The mean and standard deviation in marks of male students are 54.3 and 17.5 respectively, whereas those of female students are 50.6 and 18.0. Is there a significant difference in marks of male and female students ?

$$\left[Ans:\ \frac{\text{Difference}}{\text{S. E.}}=\frac{3.7}{1.16}=3.2\right]$$

PROBLEM 15.42. Suppose it has been found out that the first quartile height of 580 post-graduate students of a college is 165.71 cms. and the third quartile height for the same is 169.21 cms. with a standard deviation of 10.32 cms. Find out the standard errors of the two quartiles.

[*Ans* : s.e.q.=.5835]

PROBLEM 15.43. A random sample of 200 villages was taken from a district and the average population per village was found to be 485 with a standard deviation of 50. Another random sample of 200 villages from the same district gave an average population of 510 per village with a standard deviation of 40. Is the difference between the average of two samples statistically significant ?

[*Ans* : $s.e._{m_1-m_2}=4.53$]

PROBLEM 15.44. A random sample of 200 villages in a district gives the mean population of 500 persons per village. Another sample of 300 villages from the same district gives the mean at 504. If the standard deviation of the mean population of villages in that district is 20, find out if the mean of the first sample is significantly different from the combined means of the two samples taken together.

[*Ans* : $s.e._{m_1-m_2}=1.1$]

PROBLEM 15.45. Random samples drawn from two countries gave the following data relating to the height of adult males :

	Country x	*Country y*
Mean Height (in cms.)	167.42	167.25
Standard deviation	6.80	6.70
Number of observations	1000	1200

Is the difference between the mean heights significant ?

[*Ans* : $s.e._{m_1-m_2}=.29$]

PROBLEM 15.46. The mean lifetime of electric bulbs produced by a company has in the past been 1,120 hours with a standard deviation of 125 hours. A sample of 8 electric bulbs recently chosen from a supply of newly produced bulbs showed a mean lifetime of 1,070 hours. Test the hypothesis that the mean lifetime of the bulbs has not changed.

$$\left[Ans:\ \frac{\text{Difference}}{\text{S. E.}}=\frac{50}{40.2}=1.24\right]$$

Sampling of Variables— Large Samples (contd.)

16

This chapter deals with significance of difference between sample means, medians and standard deviations. To test the significance of the difference, standard error of the difference is calculated and the actual difference is compared with it. If the actual difference exceeds thrice its standard error, it is significant otherwise not. The following formula have been used in this chapter :—

(1) *Standard Error of the difference of Means: Unrelated Series.*

(*a*) Where population σ is not known $=\sqrt{\frac{\sigma^2_1}{n_1}+\frac{\sigma_2^2}{n_2}}$

(*b*) Where population σ is known $=\sqrt{\sigma^2 p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

(*c*) Where a sample means is compared with $=\sqrt{\sigma^2 p\,\frac{n}{n_1(n_1+n_2)}}$ the combined mean of 2 samples.

Related Series.

Standard error of the difference of two means. $=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{\sigma_2}-2r\frac{\sigma_1\times\sigma_2}{n_1\times n_2}}$

(2) *Standard Error of the difference of two sample medians.*

$$=\sqrt{s.e.^2 mdn_1 - s.e.^2 mdn_2}$$

(3) Standard Error of the difference of two standard deviations.

(*a*) Where population standard deviation is not known. $=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{2n_2}}$

(*b*) ,, ,, ,, ,, ,, ,, known. $=\sqrt{\frac{\sigma^2 p}{2}\left\{\frac{1}{n_1}+\frac{1}{n_2}\right\}}$

(*c*) Where S. D. of a sample is compared with the combined S. D. of both the samples $\sqrt{\frac{\sigma^2}{2}\times\frac{n}{n_1(n_1+n_2)}}$

PROBLEM 16.1. *What is meant by the standard error and what are its practical uses ? Intelligence tests of two groups of boys and girls give the following results. Examine „ the difference is significant :—*

Girls Mean 84 : S. D. 10 ; No. 121.

Boys Mean 81 : S. D. 12 ; No. 81.

SOLUTION. Supposing that the two samples are independent and come from different universes under simple sampling conditions, the standard error of the difference of their mean scores would be

$$s.e.\, m_1 - m_2 = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Substituting the given values, we get

$$s.e.\, m_1-m_2=\sqrt{\frac{(10)^2}{121}+\frac{(12)^2}{81}}=\sqrt{2.6042}=1.61.$$

The observed difference is 84—81, or 3, which is only 1.86 times this standard error and so could have arisen through the fluctuations of simple sampling and as such the difference observed in the mean scores of two intelligence tests is not significant.

PROBLEM 16.2. *The intelligence quotient (I. Q.) ratings of a group of 68 left-handed students showed a mean of 110 .52. A similar group of 68 right-handed students had a mean I. Q. of 109.48 Are the left-handed students actually above the right-handed in respect of I. Q. rating or is the difference so slight as perhaps to have been due to chance variations arising in sampling ? For left-handed students* $N_1=68$; $x_1=110.52$; $\sigma_1=15.2$. *For right-handed students* $N_2=68$; $x_2=109.48$; $\sigma_2=15.5$;,

SOLUTION. Supposing that two samples are drawn quite independently from different universes, the standard error of the difference of the two mean I. Qs. is

$$s.e.\, m_1-m_2=\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}$$

Substituting the given values. we get

$$s.e.\, m_1-m_2=\sqrt{\frac{(15.1)^2}{68}+\frac{(15.5)^2}{68}}=\sqrt{6.9}=2.6$$

The difference observed between the mean I. Q's. is 1.04, which is even less than half of this standard error and so most probably would have arisen through the fluctuation of simple sampling.

So, it would be wrong to conclude that the left-handed students are actually above the right handed in respect of I Q. rating.

PROBLEM 16.3. *From the data given below, find out whether there is a real difference in favour of the soldiers of Scotch extraction in matter of weight, or is the difference so slight that it may be attributable to chance ?*

	N	*X*	σ
Soldiers of Scotch extraction	*1821;*	*144.93 pounds;*	*17.41 pounds.*
Soldiers of French extraction	*746;*	*142.16 pounds;*	*16 .04 pounds*

SOLUTION. Since the two samples are drawn quite independently from different universes, the standard error of the difference of the two mean weights is

$$s.e.\, m_1-m_2=\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}$$

Substituting the given values, we get,

$$s.e.\, m_1-m_2=\sqrt{\frac{(17{,}41)^2}{1821}+\frac{(16.04)^2}{746}}=\sqrt{.5114}=.715$$

The observed difference between the mean weights of soldiers of Scotch extraction and French extraction is 2.77 lbs. which is more than three times this standard error. Hence the difference in weight is a real one in favour of the soldier of Scotch extraction.

PROBLEM 16.4. *The heights of a random sample of 1304 Scotchmen had a mean of 168.5456 cms. and a standard deviation of 5.480 cms. A random sample of 1694 Englishmen gave an average height of 167.4375 cms., and a standard deviation of 5.548 cms. Is the difference significant ?*

SOLUTION. The standard error of the difference of the two means given above is

$$s.e._{m_1-m_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Substituting the values, we get,

$$s.e._{m_1-m_2}=\sqrt{\frac{(5.48)^2}{1304}+\frac{(5.548)^2}{6194}}=\sqrt{.028}=.17$$

The observed difference between the two means (168.5456—167.4375) is equal to 1.1081 cms. which is more than three times the standard error of the difference of means and as such is highly significant.

PROBLEM 16.5. *A random sample of 200 villages was taken from Gorakhpur district and the average population per village was found to be 485 with a standard deviation of 50. Another random sample of 200 villages from the same district gave an average population of 510 per village with a standard deviation of 40. Is the difference between the averages of the two samples statistically significant? Give reasons.*

SOLUTION. Supposing that the two samples have been drawn quite independently the standard error of the difference of the two means is

$$s.e._{m_1-m_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Substituting the given values, we get,

$$s.e._{m_1-m_2}=\sqrt{\frac{(50)^2}{200}+\frac{(40)^2}{200}}=\sqrt{20.5}=4.53$$

The observed difference of the two average population per village is 510—485, or 25, which is thus 5.5 times this standard error and so could not have arisen through the fluctuations of simple sampling. Hence, the difference between the averages of the two samples is statistically significant.

PROBLEM 16.6. *A random sample of 1000 men from Northern India shows their mean wage to be Rs. 5 per day with a standard deviation of Rs. 1.50. A sample of 1500 men from Southern India given a mean wage of Rs. 4.50 per day with a standard deviation of Rs. 2. Discuss the suggestion that the mean rate of wages varies as between the two regions.*

SOLUTION. Supposing that the two samples have been drawn quite independently from different universes, the standard error of the difference of their mean wages is

$$s.e._{m_1-m_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Substituting the given values, we get

$$s.e._{m_1-m_2}=\sqrt{\frac{(1.5)^2}{1000}+\frac{(2)^2}{1500}}=\sqrt{.0049}=.07$$

The observed difference of the two mean wages is 0.5 which is more than three times this standard error, and so could not have arisen through the fluctuations of simple sampling. Hence, the suggestion that the mean rate of wages varies as between Northern and Southern India is correct.

PROBLEM 16.7. *A random sample of 1000 farms in a certain year gives an average yield of barley of 2000 lbs. per acre, with a standard deviation of 192 lbs. A random sample of 1000 farms in the following year gives an average yield of 2100 lbs. per acre with a standard deviation of 224 lbs. Show that these data are inconsistent with the hypothesis that the average yields in the country as a whole were the same in two years.*

SOLUTION. Supposing that the two samples have been drawn quite independently, the standard error of the difference of their mean yields is

$$s.e._{m_1-m_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Substituting the given values, we get

$$s.e._{m1-m2}=\sqrt{\frac{(192)^2}{1000}+\frac{(224)^2}{1000}}=\sqrt{87.40}=9.33.$$

The observed difference of 100 lbs. per acre which is thus more than three times this standard error and so could not have arisen through the fluctuations of simple sampling. Hence, the data are inconsistent with the hypothesis that the average yields in the country as a whole were the same in the two years.

PROBLEM 16.8. *The average chest girth in 74,459 males classed as grade I was found to be 90.8 cms. with a standard deviation of 1.94 cms. For 2146 males classed as grade II the average girth was 85.8 cms. with a standard deviation of 2.01 cms. Is there a significant difference in the chest girth of the two grades ?*

SOLUTION. Supposing that two samples have been drawn independently from different universes, the standard error of the difference of two means is

$$s.e._{m1-m2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Substituting the values, we get,

$$s.e._{m1-m2}=\sqrt{\frac{(1.94)^2}{74,459}+\frac{(2.01)^2}{2146}}=.044$$

The observed difference of 5 cms. is more than hundred times the standard error of the difference and as such is highly significant.

PROBLEM 16.9. *If 60 new entrants in a given university are found to have a mean height of 168.60. cms. and 50 seniors a mean height of 169.51 cms; is the evidence conclusive that the mean height of the seniors is greater than that of the new entrants ? Assume the standard deviation of height to be 2.48 cms.*

SOLUTION. Since the two samples are independent and come from the same universe under simple sampling conditions, the standard error of the difference of two mean heights is,

$$s.e._{m1-m2}=\sqrt{\sigma^2p\left(\frac{1}{n_2}+\frac{1}{n_2}\right)}$$

Substituting the given values, we get,

$$s.e._{m1-m2}=\sqrt{(2.48)^2\left(\frac{1}{60}+\frac{1}{50}\right)}=.47$$

The observed difference of the two mean heights is .91 cms. which is only two times this standard error and so could have arisen through the fluctuations of simple sampling. Hence, it cannot be said that the mean height of the seniors is greater than that of the new entrants.

PROBLEM 16.10. *The mean produce of wheat of a sample of 100 fields comes to 200 lbs. per acre with a standard deviation of 10 lbs. Another sample of 150 fields gives the mean at 220 lbs. with a standard deviation of 12 lbs. Assuming the standard deviation of the yield at 11 lbs. for the universe find out if there is a significant difference between the mean yields of the two samples.*

SOLUTION. Supposing the samples are independent and come from the same universe, the standard error of the difference of the mean yields would be

$$s.e._{m1-m2}=\sqrt{\sigma^2p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Substituting the given values, we get,

$$=\sqrt{(11)^2\left(\frac{1}{100}+\frac{1}{150}\right)}=1.42$$

The observed difference between the two means is 20 lbs. which is more an thrice the standard error of the difference of means. Hence the difference s significant and could not have arisen due to fluctuations of sampling.

PROBLEM 16.11. *50 children are given a special diet for a certain period nd a control group of 50 other children have normal meals. Their gains in eight are*

Specials 7.2 lbs. Controls 5.7 lbs.

The standard deviation for the weight gain is 2 lbs. Would you conclude that he special diet really promotes weight ?

SOLUTION. Since the two samples are independent the standard error of the difference of the means is

$$s.e._{m1-m2}=\sqrt{\sigma p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Substituting the-values, we get,

$$s.e._{m1-m2}=\sqrt{(2)^2\left(\frac{1}{50}+\frac{1}{50}\right)}=.4$$

The observed difference in means is (7.2—5.7) 1.5 lbs. which is more than thrice the standard error of the difference and as such very significant. The special diet does promote weights.

PROBLEM 16.12. *A random sample of 100 villages in a district gives the mean population of 500 persons per village. Another sample of 150 villages from the same district gives the mean at 504. If the standard deviation of the mean population of villages in that district is 20, find out if the mean of first samples significantly differs from the combined mean of the two samples taken together.*

SOLUTION. The combined means of the two samples.

$$=\frac{(100\times 500)+(150\times 504)}{100+150}\text{ persons,}=502{\cdot}4\text{ persons.}$$

The difference between the first sample mean and the combined mean thus is (502.4—500)

=2.4.

The standard error of the difference of first sample mean and the combined mean is

$$=\sqrt{\sigma^2 p\,\frac{n_2}{n_1(n_1+n_2)}}$$

Substituting the values,

$$=\sqrt{(20)^2\frac{150}{100(100+150)}}=\sqrt{2.4}\;=1.549.$$

Thus, the observed difference is less than thrice the standard error of the difference and as such it could not have arisen due to fluctuations of sampling. Hence the difference is not significant.

PROBLEM 16.13. *In order to find out the average daily earnings of semi-skilled workers in the shawl manufacture trade of Srinagar, an investigator took a random sample of 350 workers, and his investigation gave the average daily earnings to be Rs 5.50. He took another random sample of 300 workers and found the average earning to be Rs. 5.75. Is there significant difference between the average earnings of the group and the average earnings of both the groups taken together, assuming that the standard deviation of the average daily earnings is 60 Ps.*

SOLUTION. The average earnings of both the groups taken together are

$$= \frac{1925+1725}{350+300} \text{ rupees} = \text{Rs. } 5.61$$

The average earnings of the first sample is 5.50

Thus, the difference between the two is Re. 0.11

Supposing that the two samples have been drawn quite independently from the same universe, the standard error of the difference between the average earnings of the first sample and the average earnings of both the samples taken together is

$$s.e._{m-m_1} = \sqrt{\sigma^2 p \; \frac{n}{n_1(n_1+n_2)}}$$

Substituting the values, we get,

$$s.e._{m-m_1} = \sqrt{(60)^2 \frac{30}{350\,(350+300)}} = \sqrt{1.58} = 1.4$$

Thus, the observed difference is less than this standard error and so could have arisen through the fluctuations of simple sampling. Hence, the difference between the two average earning is not significant.

PROBLEM 16.14. *In an intelligence test administered to 60 fathers and their 100 children, the following results were obtained.*

Fathers Mean score 114 ; S. D.=13

Sons Mean score 110 ; S. D.=11

Assuming the r between the two to be +.75, calculate the standard error of the difference of the two means and state whether the difference is significant ?

SOLUTION. Since the two samples are related, the standard error of the difference of their mean scores is

$$s.e._{m_1-m_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} - 2r \times \frac{\sigma_1}{n_1} \times \frac{\sigma_2}{n_2}}$$

Substituting the values, we get,

$$s.e._{m_1-m_2} = \sqrt{\frac{(13)^2}{60} + \frac{(11)^2}{100} - 2 \times .75 \times \frac{13}{60} \times \frac{11}{100}}$$

$$= \sqrt{3.9110} = 1.99$$

Thus, the standard error of the difference of the two mean scores is 1.99. The observed difference is 114-110, or 4, which is only twice this standard error and so could have arisen through the fluctuation of simple sampling. Hence, the difference is not significant.

PROBLEM 16.15. *The following data relate to the weight (in pounds) of 100 husbands and their 100 wives :*

	Mean	*Standard deviation*
Husbands	*125*	*8*
Wives	*110*	*6*

Assuming the r between the two to be +.6. calculate the standard error of the difference of the two means and state whether the difference is significant.

SOLUTION. Since the samples are correlated the standard error of the difference of means is

$$s.e._{m_1-m_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} - 2r \times \frac{\sigma_1}{n_1} \times \frac{\sigma}{n_2}}.$$

$$= \sqrt{\frac{(8)^2}{100} + \frac{(6)^2}{100} - 2 \times .6 \times \frac{8 \times 6}{100 \times 100}}$$

$$=\sqrt{\frac{100}{100}-\frac{1.2\times 8\times 6}{100\times 100}}=\sqrt{\frac{9942.4}{10000}}$$
$$=\sqrt{.99424}=.99.$$

The observed difference in the mean weights is more than thrice the standard error of the difference of means and as such it is significant.

PROBLEM 16.16. *The following data relate to the height measurement of 100 husbands and their 120 wives.*

	Mean Height	*Standard Deviation*
Husbands	*164*	*5*
Wives	*161*	*4*

Assuming that the co-efficient of correlation between the height of husbands and wives is + .6, find out if the difference in heights is significant.

SOLUTION. Since the two samples are related the standard error of the difference of their mean height is

$$s.e._{m_1-m_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}-2r\times\frac{\sigma_1}{n_1}\times\frac{\sigma_2}{n_2}}$$
$$=\sqrt{\frac{(5)^2}{100}+\frac{(4)^2}{120}-(2\times .6)\times\frac{5}{100}\times\frac{4}{120}}$$
$$=\sqrt{.381}=.617$$

The observed difference of 3 cms. in the mean heights of husbands and wives is more than thrice the standard error and as such is highly significant.

PROBLEM 16.17. *In a wheat variety test conducted over a wide area, the mean difference between two varieties was found to be 5.5 bushels to the acre. The standard error of this difference S was 1.4 bushels per acre, and was determined from 100 pairs of plots. Set up the fiducial limits at the 5% probability level for the mean difference in yield between the two varieties.*

SOLUTION. At 5% probability level, the fiducial limits are given by M±1.96 times the standard error.

Hence, the fiducial limits for the mean difference in yield between the two varieties at the given probability level is 5.5 bushels ± (1·96 × 1.4) bushels that is 2.756 bushels and 8.244 bushels.

PROBLEM 16.18. *Two samples of 100 and 80 students are taken with a view to find out their average monthly expenditure. It is found out that the median monthly expenditure for the first group is Rs. 185 and for the second group is Rs. 200.*

The standard deviation for the first group is Rs. 7 and for the second. Rs. 8.

Examine if the difference between the median monthly expenditure of the two samples is statistically different ?

SOLUTION. Assuming that the conditions of simple sampling hold true, the standard error of the median monthly expenditure is

$$s.e._{mdn}=1.25331\frac{\sigma}{\sqrt{N}}$$

Substituting the values we get,

$$s.e._{mdn}\text{ for the 1st sample}=\frac{1.25331\times 7}{\sqrt{100}}=.877.$$

and

$$s.e._{mdn}\text{ for the 2nd sample}=\frac{1.25331\times 8}{\sqrt{80}}=1.121$$

Supposing that the two samples have been drawn quite independently the standard error of the difference of their median monthly expenditure is

$$s.e._{mdn_1-mdn_2}=\sqrt{s.e.^2_{mdn_1}+s.e.^2_{mdn_2}}$$

Substituting the calculated values, we get,

$$s.e._{mdn_1-mdn_2}=\sqrt{(.877)^2+(1.121)^2}=\sqrt{2.0258}=1.42.$$

The observed difference of the median monthly expenditure of the two samples is Rs. 200.85, or Rs. 15, which is more than ten times this standard error and so could never have arisen through the fluctuation of sampling.

Hence the difference between the median monthly expenditure of the two samples is statistically significant.

PROBLEM 16.19. *The median height of 100 M. Com. students is 166 with a standard deviation of 3, and the median height of 121 M. A. (Economics) students is 164 with a standard deviation of 4. Are M. Com. students taller than M. A. (Economics) students ?*

SOLUTION. Assuming that simple sampling conditions hold true, standard error of the height of the samples would be—

$$s.e._{mdn}=1.25331\ \frac{\sigma}{\sqrt{N}}$$

Substituting the values, we get,

$$s.e._{mdn}\text{1st Sample}=\frac{1.25331\times 3}{\sqrt{100}}=.375993$$

$$s.e._{mdn}\text{2nd Sample}=\frac{1.25331\times 4}{\sqrt{121}}=.45575$$

Supposing that the two samples have been drawn independently the standard error of the difference of the median height would be

$$s.e._{mdn_1-mdn_2}=\sqrt{s.e.^2_{mdn_1}+s.e.^2_{mdn_2}}$$

Substituting the calculated values,

$$=\sqrt{(.376)^2+(.456)^2}$$

$$=\sqrt{.3493}$$

$$=.5$$

The observed difference in the heights is 2 cms. which is more than thrice the standard error of the median, as such the difference is significant. Hence, we can conclude that M. Com. students are taller than M. A. students.

PROBLEM 16.20. *The standard deviation of the height of 100 M. Com. students is 3 cms. and of another 200 M. A. students is 4 cms. Test the significance of the standard deviations on the assumption that the standard deviation of the height of post-graduate students is 3.5 cms.*

SOLUTION. Supposing that the two samples are independent and have come from a universe which is normal. the standard error of the difference of the standard deviation is

$$s.e.\ \sigma_1\text{-}\sigma_2=\sqrt{\frac{\sigma^2_p}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$=\sqrt{\frac{(3.5)^2}{2}\left(\frac{1}{100}+\frac{1}{200}\right)}$$

$$=\sqrt{\frac{6.125\times 3}{200}}=\sqrt{.091875}=.303$$

The observed difference in the two standard deviations is 1 cm. which is slightly more than thrice the standard error of the difference of deviations, and hence the difference is significant.

PROBLEM 16.21. *For two groups of students it was found that there was no significant difference in their mean weights. However, the standard deviation in the first group (of 100 students) was 10 lbs. and in the second group (of 150 students) was 13 lbs. Is tnere a significant difference in the two standard deviation ?*

SOLUTION. On the hypothesis that the two samples are drawn from the population with identical mean and variance the standard deviation of the universe would be

$$\sigma p = \sqrt{\frac{n_1\sigma_1{}^2 + n_2\sigma_2{}^2}{n_1 + n_2}} = \sqrt{\frac{100 \times 100) + (150 \times 169)}{100 + 100}} = \sqrt{141.4}$$

The standard error of the difference of two standard deviations

$$s.e.\ \sigma_1\text{-}\sigma_2 = \sqrt{\frac{\sigma^2 p}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{141.4}{2}\left(\frac{1}{100} + \frac{1}{150}\right)} = 1.08$$

The actual difference of 3 lbs. in the two standard deviations is less than three times the standard error and as such the difference is not significant.

PROBLEM 16.22. *To find out the mean weights of girl students two independent samples of 100 each are taken which reveal no significant difference between the means ; for the first sample the standard deviation is 2.5 cms. while for the second it is 3.5 cms. Test the significance of the difference between these two measures taking the standard deviation of the universe to be 2.85 cms.*

SOLUTION. Supposing that the two samples are independent and have come from the universe which is normal, the standard error of the difference of their standard deviation is

$$s.e.\sigma_1\text{-}\sigma_2 = \sqrt{\frac{\sigma^2}{2}\left\{\frac{1}{n_1} + \frac{1}{n_2}\right\}}$$

Substituting the values we get,

$$s.e.\sigma_1\text{-}\sigma_2 = \sqrt{\frac{(2.85)^2}{2}\cdot\left\{\frac{1}{100} + \frac{1}{100}\right\}}$$

$$= \sqrt{.081225} = .285$$

The observed difference of the standard deviations of the two samples is 1 cm. which is more than three times this standard error and so could have not arisen through the fluctuations of simple sampling.

Hence, so far as the data are concerned, the difference between these two measures is significant.

PROBLEM 16.23. *The standard deviation of the weights of 746 United States soldiers of French extraction was shown to be 16.03 pounds. At the time of demobilization, weight measurements were made out only of the United States soldiers of French extraction out of 80,000 soldiers of all types. For the entire group the value of σp was 17.06 pounds. Does the observed σ for the French differ significantly from this value ?*

SOLUTION. Supposing that the distribution is normal, the standard error of the difference of standard deviations of weight of the soldiers of the French extraction and of all types is

$$s.e.\ \sigma_0\text{-}\sigma_1 = \sqrt{\frac{\sigma^2}{2} \times \frac{n^2}{n_1(n_1 + n_2)}}$$

Substituting the values, we get,

$$s.e.\ \sigma_0\text{-}\sigma_1 = \sqrt{\frac{(17.06)^2}{2} \times \frac{79254}{746\ (746 + 79254)}} = .44$$

The observed difference of the standard deviations is 1.03 pounds which is only 2.3 times this standard error and so can be attributed to the fluctuations of sampling. Hence, the data do not reveal a significant difference between the two standard deviations.

PROBLEM 16.24. *The standard deviation of the wages of 400 textile workers is Rs. 12.8, another sample of 600 textile workers gives the standard deviation at Rs. 15.7. Find out if the standard deviation of the first sample significantly differs from the combined standard deviation of the two samples which is Rs. 14.*

SOLUTION. Supposing that the distribution is normal the standard error of the difference of the deviation of the first sample and the combined standard deviation of the two samples is

$$s.e.\ \sigma_0\text{-}\sigma_1 = \sqrt{\frac{\sigma_0^2}{2} \times \frac{n_1}{n_1\,(n_1+n_2)}}$$

Substituting the values, we get,

$$s.e.\ \sigma_0\text{-}\sigma_1 = \sqrt{\frac{(14)^2}{2} \times \frac{600}{400(400+600)}}$$

$$= \sqrt{\frac{98 \times 600}{400 \times 1000}} = \sqrt{.14} = .38.$$

The observed difference between the standard deviation of the first sample and the combined standard deviation is 2.9 which is more than thrice the standard error. Hence the difference is significant.

PROBLEM 16.25. *In a sample of 1000 the mean is found to be 17.5 and the standard deviation 2.5. In another sample of 800 the mean is 18 and standard deviation 2.7. Assuming that the samples are independent discuss whether the two samples can have come from universes which have the same standard deviation.*

SOLUTION. Supposing that the samples are drawn independently the standard error of the difference of their standard deviation is

$$s.e.\ \sigma_1\text{-}\sigma_2 = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma^2_2}{2n_2}}$$

Substituting the values. we get,

$$s.e.\ \sigma_1\text{-}\sigma_2 = \sqrt{\frac{(2.5)^2}{2 \times 1000} + \frac{(2.7)^2}{2 \times 800}} = .088$$

The observed difference of the standard deviation of the samples is 0.2 which is less than three times this standard error and so might have arisen as sampling fluctuations. Hence, the two samples may have come from universes which have the same standard deviation,

PROBLEM 16.26. *For two groups of left-handed and right- handed students it was found that no significant difference existed between the two mean I. Qs. For left-handed students,* σ_1 *was 15.2 ; while for right-handed students,* σ_2 *was 15.5. Is the difference between these measures significant ? It will be remembered that* $N_1=N_2=68$.

SOLUTION. Supposing that the samples are drawn independently, the standard error of the difference of their standard deviations is

$$s.e.\ \sigma_1\text{-}\sigma_2 = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$

Substituting the given values, we get

$$s.e.\ \sigma_1\text{-}\sigma_2 = \sqrt{\frac{(15.2)^2}{2 \times 68} + \frac{(15.5)^2}{2 \times 68}} = 1.86$$

The observed difference between the standard deviation of the two samples is 15.5—15.2=0.3. This is only .16 times the standard error and so might have arisen as sampling fluctuation. Hence the difference between these measures canot be taken as significant.

PROBLEM 16.27. *The following data relate to the wages paid to workers in certain industry.*

	1st sample	*2nd sample*
Number of workers	*586*	*648*
Mean weekly wage	*Rs. 52.5*	*Rs. 47.5*
Standard deviation of wages	*10*	*11*

Find out if the mean wage and the standard deviation of wages of the first sample significantly differ from the combined mean and the combined standard deviation of the two samples.

SOLUTION. The combined mean of the two samples is

$$m_{12}=\frac{n_1m_1+n_2m_2}{n_1+n_2}=\frac{(586\times 52.5)+(648\times 47.5)}{586+648}=49.9$$

The combined standard deviation of the two samples is,

$$\sigma_{12}=\sqrt{\frac{n_1\sigma_1{}^2+n_1\,(m_1-m_{12})^2+n_2\sigma_2{}^2+n_2\,(m_2-m_{12})^2}{n_1+n_2}}$$

$$=\sqrt{\frac{(586\times 100)+586(52.5-49.9)^2+(648\times 121)+648(47.5-49.9)^2}{586+648}}$$

$$=\sqrt{\frac{144701.84}{1234}}=10.8$$

The standard error of the difference of the first sample mean and the combined means of the two samples is

$$s.e.\,m_1\text{-}m_{12}=\sqrt{\sigma^2p\,\frac{n_2}{n_1(n_1+n_2)}}$$

Substituting the value, we get,

$$=\sqrt{(10.8)^3\,\frac{648}{586\,(586+648)}}$$

$$=\sqrt{116.66\,\frac{648}{596\times 1234}}=\sqrt{.104}=.32$$

The observed difference (52.5—49.9) of 2.6 rupees is more than thrice the standard error and hence is significant.

The standard error of the difference of the standard deviation of the first sample and the combined standard deviation of the two samples is

$$s.e.\,\sigma_1\text{-}\sigma_{12}=\sqrt{\frac{\sigma^2}{2}\times\frac{n_2}{n_1\,(n_1+n_2)}}$$

Substituting the values, we get,

$$=\sqrt{\frac{(10.8)^2}{2}\times\frac{648}{586(586+648)}}=\sqrt{.052}=.23$$

The observed difference (10—10.8) of .8 is more than thrice the standard error and hence is significant. Thus both the mean and standard deviation of the first sample significantly differ from the combined mean and combined standard deviation of the two samples.

REVISIONAL PROBLEMS -16

PROBLEM 16.28. A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Show that the S. E. of the proportion of bad pineapples in a sample of this size is 0.015 and deduce that the percentage of bad pineapples in the consignment almost certainly lies between 8.5 and 17.5.

(Ans: S.E.P = 0.015 or 1.5%)

PROBLEM 16.29. The mean of a random sample of 100 individuals from a population is 64.3. The standard deviation of the sample is 2.7. Would it be unreasonable to suppose that the mean of the population is 60.

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S.E.}} = 15.93, \text{ Hypothesis is doubtful}\right)$$

PROBLEM 16.30. A railway Company installed two sets of 50 Burma ties each. The two sets were coated with creosote by two different processes. After a number of years of service it was found that 22 ties of first set and 18 ties of the second set were still in good condition. Are we justified in claiming that there is no real difference between the preserving properties of the two processes ?

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S. E.}} = \frac{.08}{.98} = .82 \text{ Not at all significant}\right)$$

PROBLEM 16.31. A random sample of 400 members is found to have a mean of 4.45 cm. Can it be reasonably regarded as a sample from a large population whose mean is 5 cm. and whose variance is 4 ?

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S.E.}} = \frac{.55}{.1} = 5.5 \text{ No.}\right)$$

PROBLEM 16.32. 500 articles were selected at random out of a batch containing 10000 articles and 30 were found to be defective. How many defective articles would you reasonably expect to find in the whole batch ?

(Ans: Between 282 and 918)

PROBLEM 16.33. In a recent survey, two samples were drawn each containing 500 hamlets. In the first sample the mean population per hamlet was found to be 100 with a S.D. of 20, while in the second sample mean population was 120 with S.D. of 15. Do you find the average of the two samples to be statistically significant ?

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S.E.}} = 18.2, \text{ Difference is significant}\right)$$

PROBLEM 16.34. Two groups consisting of 400 and 500 persons have mean heights 168.5 cms. and 166.1 cms. and variances 6.4 and 6.0 respectively. Examine whether the difference in means is significant.

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S.E.}} = \frac{2.14}{.17} = 12.3. \text{ Difference is significant}\right)$$

PROBLEM 16.35. A sample of 50 pieces of a certain type of string was tested The mean breaking strength turned out to be 14.5 lbs. Test whether the sample is from a batch of strings having a mean breaking strength of 15.6 lbs. and standard deviation of 2.2 lbs.

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S.E.}} = 3.55. \text{ Hypothesis is not correct, i.e., the sample came from a different population}\right)$$

PROBLEM 16.36. A sample of 600 persons selected randomly from a large city gives the result that the males are 53%. Is there reason to doubt the hypothesis that males and females are in equal numbers in the city.

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S.E.}} = \frac{18}{12.25} = 1.47, \text{ Hypothesis is true}\right)$$

PROBLEM 16.37. A random sample of 100 students gave a mean weight of 58 kilograms with S.D. of 4 kg. Test the hypothesis that the mean weight in the population is 60 k.g.

$$\left(\text{Ans: } \frac{\text{Difference}}{\text{S.E.}} = \frac{2}{4} = 0.5\right)$$

X^2 (Chi-Square) Test and Co-efficient of Contingency

17

X^2 test is a test of significance. It is a very useful device of checking observed frequencies against expected frequencies. It is a measure between a sample and an accepted theory or standard. The value of X^2 can be calculated by means of the following formulae, which are explained in detail in the solution of different problems :

(*i*) $X^2 = S\left\{\left(\frac{f-f_1}{f}\right)^2\right\}$ This formula can be applied in all types of distributions.

(*ii*) In a 2×2 table $X^2 = \frac{(b_1c_2 - c_1b_2)\,T}{T_1.\,T_2.\,Tb.\,Tc.}$

(*iii*) In a 2×*n* table $X^2 = \frac{T^2}{Tb.\,Tc.}\left\{\Sigma\left(\frac{b^2}{Ts}\right) - \frac{Tb^2}{T}\right\}$

(*iv*) where *p* and *q* are given $X^2 = \frac{\left(a - \frac{p}{q}b\right)}{\frac{p}{q}N}$

The calculated value of X^2 is compared with the table value for given degrees of freedom and at a certain level of significance. If the table value is less than the calculated value the divergence between observed and expected frequencies is considered to be significant.

In this chapter, problems on co-efficient of contingency have also been solved. If statistical data are divided by dichotomy by observing only two attributes, say A and B it is possible to study the relationship between them by the calculation of co-efficient of association. But if A is further sub-divided in various classes and similarly if B is also sub-divided, the study of association becomes difficult. For such tables (which are known as contingency tables) co-efficient of contingency has to be calculated. Karl Pearson has given the following formula for the calculation of co-efficient of contingency :—

$$C = \sqrt{\frac{X^2}{N + X^2}}$$

In a	2× 2 table the Maximum Value of C	=.707
,,	3× 3 ,, ,, ,, ,,	=.816
,,	4× 4 ,, ,, ,, ,,	=.866
,,	5× 5 ,, ,, ,, ,,	=.894
,,	6× 6 ,, ,, ,, ,,	=.913
,,	7× 7 ,, ,, ,, ,,	=·926
,,	8× 8 ,, ,, ,, ,,	=.935
,,	9× 9 ,, ,, ,, ,,	=.943
,,	10×10 ,, ,, ,, ,,	=.949

PROBLEM 17.1. *Population samples from six small frontier mining camps are given below.*

	Camps					
	1	*2*	*3*	*4*	*5*	*6*
Males	*22*	*12*	*32*	*18*	*46*	*28*
Females	*6*	*8*	*12*	*4*	*24*	*22*

Compute the value of X^2

SOLUTION. Let us take as our hypothesis the supposition that males and females are equally distributed in camps. If this be true, the expected frequencies would be

	Camps					
	1	2	3	4	5	6
Males	14	10	22	11	35	25
Females	14	10	22	11	35	25

Substituting the observed and the theoretical frequencies in the formula

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\};$$

Where f represents the observed frequencies, f_1 the expected or theoretical frequencies and S, the summation.

We get,

$$X^2 = \frac{(22-14)^2}{14} + \frac{(6-14)^2}{14} + \frac{(12-10)^2}{10} + \frac{(8-10)^2}{10}$$
$$+ \frac{(32-22)^2}{22} + \frac{(12-22)^2}{22} + \frac{(18-11)^9}{11} + \frac{(4-11)^2}{11}$$
$$+ \frac{(46-35)^2}{35} + \frac{(24-35)^2}{35} + \frac{(28-25)^2}{25} + \frac{(22-25)^2}{25}$$

or $$X_2 = \frac{64}{14} + \frac{64}{14} + \frac{4}{10} + \frac{4}{10} + \frac{100}{22} + \frac{100}{22} + \frac{49}{11} + \frac{49}{11}$$
$$+ \frac{121}{35} + \frac{121}{35} + \frac{9}{25} + \frac{9}{25}$$

or $X^2 = 35.56.$

PROBLEM 17.2. *The following table shows the condition of home and the condition of child.*

Condition of child	*Condition of home*		
	Clean	*Not clean*	*Total*
Clean	*76*	*43*	*119*
Fairly clean	*38*	*17*	*55*
Dirty	*25*	*47*	*72*
Total	*139*	*107*	*246*

Do these results suggest that the condition of the home has a bearing on the condition of the child.

(At 5% level of significance, the value of X^2 *is 5.991 for 2 degrees of freedom.)*

SOLUTION. Let us take as our hypothesis the supposition that the condition of the home and the condition of the child are not associated, that is, the two attributes in the given table are independent. If this be true the theoretical frequencies would be :—

Condition of child	Condition of home		
	Clean	Not-clean	Total
Clean	67	52	119
Fairly clean	31	24	55
Dirty	41	31	72
Total	139	107	246

Substituting the observed and the theoretical frequencies in the formula :

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(76-67)^2}{67} + \frac{(43-52)^2}{52} + \frac{(38-31)^2}{31} + \frac{(17-24)^2}{24} + \frac{(25-41)^2}{41} + \frac{(47-31)^2}{31}$$

or $$X^2 = \frac{81}{67} + \frac{81}{52} + \frac{49}{31} + \frac{49}{24} + \frac{256}{41} + \frac{256}{31}$$

$X^2 = 20.9$ approx.

The number of degrees of freedom

$= (c-1)\,(r-1) = (2-1)\,(3-1) = 2.$

The table value of X^2 for 2 degrees of freedom is 5.991, which is much less than the calculated value, 20.9, Therefore, the hypothesis is not proved by the data. In other words, the X^2 test shows that the condition of the home has a bearing on the condition of the child.

PROBLEM 17.3. *Show that there is a significant association between the temperament of brothers and sisters.*

Good-Natured Brothers	*Sisters*			
	Quick	*Good-Natured*	*Sullen*	*Total*
Quick	*850*	*571*	*580*	*2001*
Good-Natured	*618*	*593*	*455*	*1666*
Sullen	*540*	*456*	*457*	*1453*
Total	*2008*	*1620*	*1492*	*5120*

SOLUTION. Let us take for our hypothesis the supposition that the temperament of brothers and sisters are not associated, *i.e.*, the two attributes under consideration are independent. If this be so, the expected frequencies would be :

Brothers	Sisters			
	Quick	Good-Natured	Sullen	Total
Quick	785	633	583	2001
Good-Natured	653	527	486	1666
Sullen	570	460	423	1453
Total	2008	1620	1492	5120

Substituting the observed and the theoretical frequencies in the formula :

$$X^2 = S\left\{ \frac{f-f_1{}^2}{f_1} \right\}$$

We get,

$$X^2 = \frac{(850-785)^2}{785} + \frac{(571-633)^2}{633} + \frac{(580-583)^2}{583}$$

$$+ \frac{(618-653)^2}{653} + \frac{(593-527)^2}{527} + \frac{(455-486)^2}{486}$$

$$+ \frac{(540-570)^2}{570} + \frac{(456-460)^2}{460} + \frac{(457-423)^2}{423}$$

or $$X^2 = \frac{4225}{785} + \frac{3844}{633} + \frac{9}{583} + \frac{1225}{653}$$

$$+ \frac{4356}{527} + \frac{961}{486} + \frac{900}{570} + \frac{16}{460} + \frac{1156}{423}$$

or $$X^2 = 5.38 + 6.07 + .02 + 1.88 + 8.27 + 1.98$$

$$+ 1.58 + .03 + 2.73$$

or $$X^2 = 27.94$$

The number of degrees of freedom in the given problem

$= (c-1)\,(r-1) = (3-1)\,(3-1) = 4.$

The table value of X^2 for 4 degrees of freedom at 5% level of significance is 9.488, which is much less than the calculated value of X^2 which is 27.94. This leads us to the conclusion that the hypothesis is incorrect and there is a significant association between the temperament of brothers and sisters.

PROBLEM 17.4. *In a recent diet survey the following results were found in an Indian city :—*

Number of	*Hindus*	*Muslims*	*Total*
Families taking tea	*1236*	*164*	*1400*
Families not taking tea	*564*	*36*	*600*
Total	*1800*	*200*	*2000*

Discuss whether there is any significant difference between the two communities in the matter of tea taking

SOLUTION. Let our hypothesis be that there is no difference between the two communities in the matter of taking tea, that is the two attributes are independent. If this be so, the expected frequencies would be :—

	Hindus	Muslims
Families taking tea	1260	140
Families not taking tea	540	60

Substituting the observed and the expected frequencies in the formula.

$$X^2 = S\left\{ \frac{(f - f_1)^2}{f_1} \right\}$$

We get,

$$X^2 = \frac{(1236-1260)^2}{1260} + \frac{(164-140)^2}{140} + \frac{(564-540)^2}{540} + \frac{(36-60)^2}{60}$$

$$= (24)^2 \left\{ \frac{1}{1260} + \frac{1}{140} + \frac{1}{540} + \frac{1}{60} \right\} = 15.22.$$

The number of degrees of freedom

$= (c-1)\,(r-1) = (2-1)\,(2-1) = 1.$

The value of X^2 for one degree of freedom at 5% level of significance is 3.841 which is much less than this calculated value of X^2. In the light of this the conclusion is clear that the hypothesis is wrong and there is a significant difference between the two communities in the matter of taking tea.

PROBLEM 17.5. *Suppose that, in a public opinion survey answers to the questions—*

(a) Do you drink ?

(b) Are you in favour of local option on sale of liquor

were as tabulated below :

Question (b)	*Question (a)*		
	Yes	*No*	*Total*
Yes	*56*	*31*	*87*
No	*18*	*6*	*24*
Total	*74*	*37*	*111*

Can you infer that opinion on local option is dependent on whether or not an individual drinks ?

Values of X^2 on levels of significance P

Degrees of freedom	*P=0.05*	*P=0.10*
1	*3.84*	*2.71*
2	*5.99*	*4.60*
3	*7.82*	*6.25*

SOLUTION. Let us take as our hypothesis the assumption that local option on sale of liquor and whether an individual drinks, are not associated, that is, these two attributes in the given table are independent. If this be so, the theoretical frequencies would be :—

Question (b)	Question (a)		
	Yes	No	Total
Yes	58	29	87
No	16	8	24
Total	74	37	111

Substituting the observed theoretical frequencies in the formula

$$X^2 = S\left\{ \frac{(f - f_1)^2}{f_1} \right\}$$

We get,

$$X^2 = \frac{(56-58)^2}{58} + \frac{(31-29)^2}{29} + \frac{(18-16)^2}{16} + \frac{(6-8)^2}{8}$$

$$= (2)^2 \left\{ \frac{1}{58} + \frac{1}{29} + \frac{1}{16} + \frac{1}{8} \right\} = .957$$

The number of degrees of freedom is

$$=(c-1)(r-1)=(2-1)(2-1)=1$$

The five percent value of X^2 for one degree of freedom is 3.841, which more than the calculated value of X^2. So in the light of this the conclusion evident that our hypothesis is correct and opinion on local option on sale liquor is not dependent on whether an individual drinks or not.

PROBLEM 17.6. *The table given below shows the data obtained during epidemic of cholera.*

	Attacked	*Not Attacked*	*Total*
Inoculated	*31*	*469*	*500*
Not Inoculated	*185*	*1315*	*1500*
Total	*216*	*1784*	*2000*

Test the effectiveness of inoculation in preventing the attack of cholera.

(Five percent value of X^2 for one degree of freedom is 3 .841)

SOLUTION. Let us take as our hypothesis the supposition that inoculation and absence of attack from cholera are not associated, that is, these two attributes in the given table are independent. If this be so, the theoretical frequencies would be :

	Attacked	Not Attacked	Total
Inoculated	54	436	500
Not Inoculated	162	1338	1500
Total	216	1784	2000

Substituting the observed and the theoretical frequencies in the formula

$$X^2 = S \left\{ \frac{(f-f_1)^2}{f_1} \right\};$$

we get,

$$X^3 = \frac{(31-54)^2}{54} + \frac{(469-446)^2}{446} + \frac{(185-162)^2}{162} + \frac{(1315-1338)^2}{1338}$$

$$=(23)^2 \left\{ \frac{1}{54} + \frac{1}{446} + \frac{1}{162} + \frac{1}{1338} \right\}$$

$$=14.64.$$

The number of degrees of freedom

$$=(c-1)(r-1)=(2-1)(2-1)=1.$$

The five per cent value of X^2 for one degree of freedom is 3.841, which is much less than the calculated value of X^2. So in the light of this, conclusion is evident that the hypothesis is incorrect and inoculation and absence of attack from cholera are associated.

PROBLEM 17.7. *The following table is published in a memoir written by Karl Pearson :—*

		Eye colour in sons	
		not light	*light*
Eye colour in father	*not light*	*230*	*148*
	light	*151*	*471*

Test whether the colour of the son's eyes is associated with that of the father's, (you may use the fact that 5% value of chi-square for 1 degree of freedom is 3.84)

SOLUTION. Let us take for our hypothesis, the supposition that the eye colour of sons and the eye colour of fathers are independent. If this be true, the expected frequencies would be : —

		Eye colour in sons	
		not light	light
Eye colour in fathers	not light	144	234
	light	237	385

Substituting the observed and the theoretical frequencies in the formula.

$$X^2 = S\left\{ \frac{(f-f_1)^2}{f_1} \right\}$$

We get,

$$X^2 = \frac{(230-144)^2}{144} + \frac{(148-234)^2}{234} + \frac{(151-237)^2}{237} + \frac{(471-385)^2}{385}$$

$$= (86)^2 \left\{ \frac{1}{144} + \frac{1}{234} + \frac{1}{237} + \frac{1}{385} \right\}$$

$$= 133.29.$$

The number of degrees of freedom.

$= (c-1)(r-1) = (2-1)(2-1) = 1.$

The value of X^2 at 5% level of significance for 1 degree of freedom is 3.841 which stands very poor in comparison to the calculated value, 133.29. This leads us to the definite conclusion that the hypothesis is wrong and the colour of sons' eyes is associated with that of the fathers to a great extent.

PROBLEM 17.8. *Two treatments A and B were tried to control a certain type of plant disease. The following results were obtained :—*

A : 200 plants were examined and 40 were found infected.

B : 200 plants were examined and 10 found infected.

Is treatment B superior to treatment A ?

SOLUTION. The data given in the question can be tabulated as follows :

	Infected	Not infected	Total
A	40	160	200
B	10	190	200
Total	50	350	400

On the hypothesis that there is no difference between the two treatments the expected figures would be

25	175	200
25	175	200
50	350	400

$$\text{the value of } X^2 = \frac{(40-25)^2}{25} + \frac{(10-25)^2}{25} \times \frac{(160-175)^2}{175} + \frac{(190-175)^2}{175} = 20.6$$

For one degree of freedom at 5% level of significance the value of $X^2 = 3.841$ and as the calculated value is much higher than this the results are significant and the hypothesis has to be rejected. Treatment B is superior to treatment A.

PROBLEM 17.9. *The following table shows fractional test meal results in a series of pathologically verified cases of ulcer and cancer of stomach.*

	Achlor-hydria	*Hypo-chlorhydria*	*Normal*	*Hyper-chlorhydria*	*Total*
Chronic ulcer	*3*	*7*	*35*	*9*	*54*
Cancer	*22*	*2*	*6*	*0*	*30*
Total	*25*	*9*	*41*	*9*	*84*

Is there a significant difference in acidity condition between chronic ulcer and cancer ?

SOLUTION. On the supposition that there is no significant difference in acidity conditions between ulcer and cancer, the expected frequencies would be as follows :

	a	*b*	*c*	*d*	Total
A	16	5.8	26.4	5.8	54
B	9	3.2	14.6	3.2	30
Total	25	9	41	9	84

$$\text{The value of } X^2 = \frac{(3-16)^2}{16} + \frac{(22-9)^2}{9} + \frac{(7-5.8)^2}{5.8}$$
$$+ \frac{(2-3.2)^2}{3.2} + \frac{(35-26.4)^2}{26.4} + \frac{(6-14.6)^2}{14.6}$$
$$+ \frac{(9-5.8)^2}{5.8} + \frac{(0-3.2)^2}{3.2} = 42.93$$

Number of degrees of freedom

$= (c-1)(r-1) = (4-1)(2-1) = 3.$

For 3 degrees of freedom at 5% level of significance the value of $X^2 = 7.82$ and as the calculated value is much higher than this and the difference is significant the hypothesis should be rejected.

PROBLEM 17.10. *In an experimentation the immunization of goats from anthrax the following results were obtained. Derive your inference on the efficacy of the vaccine.*

	Died of anthrax	*Survived*	*Total*
Inoculated with vaccine	*2*	*10*	*12*
Not inoculated	*6*	*6*	*12*

SOLUTION. (1) We take for our hypothesis, the supposition that inoculation and survival are independent attributes. If this be true, the expected frequencies would be :—

	Died of anthrax	Survived
Inoculated with vaccine	4	8
Not Inoculated	4	8

Substituting the observed and the expected frequencies in the formula

$$X^2 = S\left\{ \frac{(f-f_1)^2}{f_1} \right\}$$

We get,

$$X^2 = \frac{(2-4)^2}{4} + \frac{(10-8)^2}{8} + \frac{(6-4)^2}{4} + \frac{(6-8)^2}{8}$$
$$= (2)^2 \left\{ \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{4} + \tfrac{1}{8} \right\} = 3.$$

The number of degrees of freedom is

$= (c-1)(r-1); = (2-1)(2-1) = 1.$

The value of X^2 at 5% level of significance for one degree of freedom is 3.841. The calculated value which is 3 is thus less than it. Hence there is no cause to suspect the hypothesis and the data do not suggest that the survival is associated with inoculation. Whatever association there appears to be between these two from a direct comparison may be due to fluctuations of sampling.

SOLUTION. (II) There is a simple formula for calculating X for this type of table, *i. e.*, (2×2) fold table given by Brandt and Snedecor.

Representing the (2×2) fold table as follows :—

$$\begin{array}{ccc} b_1 & b_2 & T_b \\ c_1 & c_1 & T_c \\ T_1 & T_2 & T \end{array}$$

$$X^2 \text{ is given by } = \frac{(b_1c_2 - c_1b_2)^2 T}{T_1 . T_2 . T_b . T_c .}$$

We multiply diagonal frequencies and find the difference between the two products. The difference is squared and multiplied by the grand total and the result is divided by the product of the subtotals.

In our case, the table is as follows :

2	10	12
6	6	12
8	16	24

Substituting these values, we get

$$X^2 = \frac{\{(2\times 6)-(6\times 10)\}^2\times 24}{8\times 16\times 12\times 12} = \frac{48\times 48\times 24}{8\times 16\times 12\times 12} = 3.$$

PROBLEM 17.11. *Explain the technique of the X^2 test of goodness of fi, pointing out the precautions to be observed in its applications.*

Two sample polls of votes for two candidates A and B for a public office are taken, one from among residents of urban areas, and the other from residents of rural areas. The results are given below. Examine whether the nature of the area is related to voting preference in this election.

Votes for / *Area*	*A*	*B*	*Total*
Rural	*620*	*380*	*1,000*
Urban	*550*	*450*	*1,000*
Total	*1,170*	*830*	*2,000*

SOLUTION. Proceeding on the hypothesis that the two attributes—nature of area and the voting preference given in the table are not associated we get the following theoretical frequencies.

Votes for / Area	A	B	Total
Rural	585	415	1,000
Urban	585	415	1,000
Total	1,170	830	2,000

Substituting the observed and expected frequencies in the formula

$$X^2 = S\left[\frac{(f-f_1)^2}{f_1}\right]$$

We get,

$$X^2 = \frac{(620-585)^2}{585} + \frac{(380-415)^2}{415} + \frac{(550-585)^2}{585} + \frac{(450-415)^2}{415}$$

$$X^2 = \frac{(35)^2}{585} + \frac{(-35)^2}{415} + \frac{(-35)^2}{585} + \frac{(35)^2}{415}$$

$$X^2 = (35)^2\left\{\frac{1}{585} + \frac{1}{415} + \frac{1}{585} + \frac{1}{415}\right\} = 10.99$$

The number of degrees of freedom is,

$= (c-1)\,(r-1) = (2-1)\,(2-1) = 1$

The value of X^2 at 5% level of significance for one degree of freedom i 3.841. The calculated value which is 10.09 is thus more than it. The results ar gnificant and the hypothesis is incorrect. The nature of area and the voting eference are associated.

PROBLEM 17.12. *In December 1974, there was an outbreak of cholera a certain place. Of 120 persons who were uninoculated, 10 contracted the disease, 4 of them died. Of 180 persons who were inoculated 5 contracted cholera and none f them died.*

(a) *inoculation and contracting the disease, and*

(b) *inoculation and mortality among persons who have contracted the disease.*

SOLUTION. The data can be set up in the following two tables

(a)	Infected	Not infected	Total
.Uninoculated	10	110	120
Inoculated	5	175	180
Total	15	285	300

(b)	Died	Recovered	Total
Uninoculated	4	6	10
Inoculated	0	5	5
Total	4	11	15

(a) $X^2 = \dfrac{(b_1c_2 - c_1b_2)^2 \times T}{T_1 \times T_2 \times T_b \times T_c}$

$b_1 = 10$; $c_1 = 5$; $T_1 = 15$; $T_b = 120$; $T = 300$
$b_2 = 100$; $c_2 = 175$; $T_2 = 285$: $T_c = 180$;

Substituting the value, we get

$$X^2 = \frac{\{(10 \times 175) - (5 \times 110)\}^2 \times 300}{15 \times 285 \times 120 \times 180}$$

$$= \frac{(1750 - 550)^2 \times 300}{92340000} = \frac{432000000}{92340000} = 4.7$$

For one degree of freedom at 5% level of significance the table value of $X^2 = 3.841$, and as the calculated value is higher than this the results are significant and there is evidence of positive association between infection and absence of inoculation.

(b) $X^2 = \dfrac{(b_1c_2 - c_1b_2)^2 \times T}{T_1 \times T_2 \times T_b \times T_c}$

$b_1 = 4$; $c_1 = 0$; $T_1 = 4$; $T_b = 10$;
$b_2 = 6$; $c_2 = 5$; $T_2 = 11$; $T_c = 5$; $T = 15$

Substituting the values, we get

$$X^2 = \frac{\{(4 \times 5) - (0 \times 6)\}^2 \times 15}{4 \times 11 \times 10 \times 5} = \frac{6000}{2200} = 2.7$$

For one degree of freedom at 5% level of significance the table value of $X^2 = 3.841$, and as such the results are not significant. We cannot say that mortality among those who have contracted the disease is higher amongst those who were not inoculated.

In cases where the frequencies are less than 5, X^2 test should not be applied, as it gives misleading conclusions. If we calculated co-efficient of association the results would be found to be significant.

PROBLEM 17.13. *In the course of anti-malarial work in Birnagore in the third quarter of 1972, quinine was administered to 606 adults out of a total population of 3,540. The incidence of malarial fever is shown below. Discuss the preventive value of quinine.*

	Fever	*No Fever*	*Total*
Quinine ...	*19*	*587*	*606*
No Quinine ...	*193*	*2741*	*2934*
Total ...	*212*	*3328*	*3540*

You may use the 5% value of Chi-square for n degree of freedom, equal to 1, the value being 3.841.

SOLUTION. On the hypothesis that quinine has no preventive value.

The value of X^2

$$= \frac{(b_1c_2-c_1b_2)^2 T}{T_1 \times T_2 \times T_b \times T_c}$$

In the present problem

$b_1=19;\ c_1=193;\ T_1=212;\ Tb=606$
$b_2=587;\ c_2=2741;\ T_2=3328;\ Tc=2934$
$T=3540$

Substituting the values, we get,

$$X^2= \frac{(19\times2741-193\times587)^2\times3540}{212\times3328\times606\times2934}$$

$$= \frac{(52079-113291)^2\times3540}{212\times3328\times606\times2934} = \frac{(-61212)^2\times3540}{212\times3328\times606\times2934}$$

$$= \frac{A.L.\ (\log.\ 61212\times2+(\log.\ 3540)}{A.L.\ (\log.\ 212+\log.\ 3328+\log.\ 606+\log.\ 2934)}$$

$$= \frac{A.L.\ 13.1228}{A.L.\ 12.0981} = A.L.\ (13.1228-12.0981)$$

$$= A.L.\ 1.0247 = 10.59.$$

The table value of X for one degree of freedom at 5% level is 3.841. Since the calculated value is higher than the table value, the results are significant and the hypothesis has to be rejected.

PROBLEM 17.14. *The Fortune magazine of the U. S. A. published the following results of a sample survey of public opinion regarding election of Nixon as the President of U.S.A. :—*

Attitude towards election	*Rich*	*Poor*	*Total*
Favourable	*508*	*1559*	*2067*
Unfavourable	*905*	*1114*	*2019*
Total	*1413*	*2673*	*4086*

Is attitude towards election issue guided by the economic status of the voters? What test would you apply? The following table of 1% values of chi-square is reproduced for your use :—

Degrees of freedom	*1*	*2*	*3*	*4*
1% value of chi-square	*6.635*	*9.210*	*11.341*	*13.277*

SOLUTION. The value of X^2 is

$$\frac{(b_1c_2-c_1b_1)^2\times T}{T_1\times T_2\times T_b\times T_c}$$

In the present example

$b_1=508;\ c_1=905;\ T_1=1413;\ T_b=2067$
$b_2=1559:\ c_2=1114;\ T_2=2673;\ T_c=2019$
$T=4086$

Substituting the values, we get,

$$X^2=\frac{\{(508+1114)-(905\times 1559)\}^2\times 4086}{1413\times 2673\times 2067\times 2019}$$

$$=\frac{\{A.L.\ (\log.\ 508+\log.\ 1114)-A.L.\ (\log.\ 905+\log.\ 1559)\}^2\times 4086}{A.L.\ (\log.\ 1413+\log.\ 2673+\log.\ 2067+\log.\ 2019)}$$

$$=\frac{\{A.L.\ (2.7059+3.0453)-A.L.\ (2.9566+3.1931)\}^2\times 4086}{A.L.\ (3.1492+3.4265+3.3160+3.4054)}$$

$$=\frac{(A.L.\ 5.7512-A.L.\ 6.1497)^2\times 4086}{(A.L.\ 13.2971)}$$

$$=\frac{(563600-1413000)^2\times 4086}{(A.L.\ 13.9271)}$$

$$=\frac{(-849400)^2\times 4086}{(A.L.\ 13.2971)}$$

$$=\frac{A.L.\ (\log.\ 849400\times 2)+(\log.\ 4086)}{(A.L.\ 13.2971)}$$

$$=\frac{A.L.\ (11.8578+3.6117)}{A.L.\ (13.2971)}=\frac{A.L.\ (15.4695)}{A.L.\ (13.2971)}$$

$$=A.L.\ (15.4695-13.2971)=A.L.\ 2.1724.=148.7$$

The table value of X^2 for one degree of freedom at 1% level is equal to 6.635. The calculated value is much higher than this and as such the result is significant. The attitude of voters is definitely guided by the economic status.

PROBLEM 17.15. *From the following table showing the number of plants having certain characters, test the hypothesis that the flower colour is independent of flatness of leaf.*

	Flat leaves	*Curled Leaves*	*Total*
White Flowers	*99*	*36*	*135*
Red Flowers	*20*	*5*	*25*

You may use the following table giving the value of X^2 (chi-square) for one degree of freedom, for different values of P.

P—.99	.95	.90	.50	.10	.05	.01
X^2=.000157	.00393	.0158	.455	2.706	3.841	6.635

SOLUTION. On the hypothesis that the flower colour is independent of flatness of leaf the expected frequencies are :

	Flat Leaves	Curled Leaves	Total
White Flowers	100.4	34.6	135
Red Flowers	18.6	6.4	25
	119	41	160

Substituting the observed and the theoretical frequencies in the formula

$$X^2=S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2=\frac{(99-100.4)^2}{100.4}+\frac{(36-34.6)^2}{34.6}+\frac{(20-18.6)^2}{18.6}+\frac{(5-6.4)^2}{6.4}$$

$$=(1.4)^2\left\{\frac{1}{100.4}+\frac{1}{34.6}+\frac{1}{18.6}+\frac{1}{6.4}\right\}=.4878 \text{ or } .488$$

The number of degrees of freedom is

$$=(c-1)(r-1)=(2-1)(2-1)=1.$$

The 5% value of X^2 for one degree of freedom is 3.841. The calculated value is much less than it. This comparison leads us to the conclusion that there is no cause to suspect the hypothesis and the flower colour is independent of the flatness of the leaf.

PROBLEM 17.16. *In an experiment on immunization of cattle from tuberculosis, the following results were obtained :—*

	Died or affected	*Unaffected*
Inoculated	*12*	*26*
Not inoculated	*16*	*6*

Examine the effect of vaccine in controlling susceptibility to tuberculosis.

SOLUTION. Percentage of inoculated who are unaffected is $\frac{26}{38}\times 100=$ 68.4 and percentage of not inoculated who are unaffected is $\frac{6}{22}\times 100=27.3$.

Thus the comparison of the two percentages brings out clearly the fact that inoculation and non-susceptibility to tuberculosis are positively associated.

But, it is just possible that this association may be due to fluctuations of sampling. So, to test the significance of association, we proceed as follows :

Taking for our hypothesis the supposition that the two attributes given in the problem are independent, the theoretical frequencies are :

	Died or affected	Unaffected
Inoculated	17.7	20.3
Not inoculated	10.3	11.7

Substituting the observed and the theoretical frequencies in the formula

$$X^2=S\left\{\frac{(f-f_1)^2}{f_1}\right\};$$

We get

$$X^2=S\ \frac{(12-17.7)^2}{17.7}+\frac{(26-20.3)^2}{20.3}+\frac{(16-10.3)^2}{10.3}+\frac{(6-11.7)^2}{11.7}$$

$$=(5.7)^2\left\{\frac{1}{17.7}+\frac{1}{20.3}+\frac{1}{10.3}+\frac{1}{11.7}\right\}=9.367.$$

The number of degree of freedom is

$$=(c-1)(r-1)=(2-1)(2-1)=1.$$

The value of X^2 for one degree of freedom at 5% level to significance is 3.841. The calculated value is thus greater than his value. So the hypothesis is incorrect and the association between the two attributes is significant. Thus it is established that the vaccine is effective in controlling susceptibility to tuberculosis.

PROBLEM 17.17. *The following table gives the results of a series of controlled experiments. Discuss whether the treatment may be considered to have any positive effect.*

	Positive	*No effect*	*Negative*
Treatment	*9*	*2*	*1*
Control	*3*	*6*	*3*
	12	*8*	*4=24*

SOLUTION. A simple formula for calculating X^2 for $(2\times n)$ fold tables is given by Brandt and Snedecor. Representing the $(2\times n)$ fold table as follows :

$$\begin{array}{cccccc} & 1 & 2 & 3 & n & \\ b & b_1 & b_2 & b_3 & \ldots bn & T_b \\ c & c_1 & c_2 & c_3 & \ldots c_n & T_c \\ & Ts & Ts & Ts & Ts & T \end{array}$$

$$X^2 \text{ is given by } \frac{T^2}{T_b T_c}\left\{\Sigma\left(\frac{b^2}{Ts}\right)-\frac{T_b{}^2}{T}\right\}$$

"Each frequency in either of the rows is squared and divided by the corresponding sub-total. These are summated and the correction term subtracted as shown in the formula. The remainder is multiplied by the quotient of the square of the total frequency divided by the product of two sub-totals on the right" In the given example, the table is :

	Positive	No effect	Negative	
Treatment	9	2	1	12
Control	3	6	3	12
	12	8	4	24

Substituting these values in the formula, we get

$$X^2=\frac{(24)^2}{12\times 12}\left\{\frac{(9)^2}{12}+\frac{(2)^2}{8}+\frac{(1)^2}{4}-\frac{(12)^2}{24}\right\}$$

$$=\frac{24\times 24}{12\times 12}\{6.75+0.5+0.25-6.00\}=4\times 1.5=6.$$

The number of degrees of freedom is

$$(c-1)(r-1)=(2-1)(3-1)=2$$

The value of X^2 for 2 degrees of freedom at 5% level of significance is 5.99. The calculated value is almost equal to this value. Hence the treatment may be considered to have no positive effect.

PROBLEM 17.18. *The normal rate of infection for a certain disease in cattle is known to be 50%. In an experiment with given animals injected with a new vaccine it was found that none of the animals caught infection. Can the evidence be regarded as conclusive (at the 1% level of significance) to prove the value of the new vaccine ?*

SOLUTION. Representing

the chance of infection by p

the chance of non-infection by q

the number of vaccinated people who caught infection by a

the number of vaccinated people who did not catch infection by b

and the number of observations by n

we have as per given data

$p=.50$ $\quad b=7$

$q=.50$ $\quad n=a+b=7$

$a=0$

Substituting these values in the formula

$$X^2 = \frac{\left(a - \frac{p}{q}\,b\right)^2}{\frac{p}{q}\cdot n}$$

We get,

$$X^2 = \frac{\left(0 - \frac{.50}{.50}\,7\right)^2}{\frac{.50}{.50}\,7} = 7$$

The number of degrees of freedom is 1. The value of X^2 for 1 degree of freedom at 1% level of significance is 6.64 which is less than the calculated value. Hence the evidence may be regarded as conclusive to prove the value of the new vaccine.

PROBLEM 17.19. *Upon the basis of past experience the fatality rate from malaria fever for a certain community was found to be 12.5 per cent (that is, reported deaths from malaria fever reported cases of malaria fever = .1250). A survey was made of certain congested areas. The homes studied were selected as nearly as possible at random and a fatality ratio of 30.0 (45 deaths) was found for 150 cases of malaria fever. Using X^2 test, determine if it represents a significant departure from the population value 12.5.*

SOLUTION. Representing

the chance of death when attacked by malaria fever by p

tbe chance of survival when attacked by malaria fever by q

the number of deaths from malaria fever by a

the number of survivals among those who were attacked by malaria fever by b

and the number of observations by n.

we have as per data :

$p = .125$ $\quad b = 105$

$q = .875$ $\quad n = 150$

$a = 45$

Substituting these values in the formula

$$X = \frac{\left(a - \frac{p}{q}\,b\right)^2}{\frac{p}{q}\,n}$$

We get,

$$X^2 = \frac{\left(45 - \frac{.125}{.8}\,105\right)^2}{\frac{.125}{.875}\,150} = 42$$

The number of degrees of freedom is 1. The value of X^2 for a degree of freedom at 1% level of significance is 6.64 which is much less than the calculated value. Hence the sample fatality rate represents a significant departure from the population fatality rate from malaria fever.

PROBLEM 17.20. *In a certain hospital during a certain week, 50 babies were born out of whom 30 were males and 20 females. Using X^2 test, ascertain if this distribution is inconsistent with the observation that the sex ratio among births is 1:1.*

SOLUTION. Representing
the chance of a male birth by p
the chance of a female birth by q
the number of male births by a
the number of female births by b
and the total number of observations by n

we have as per data :

$p = .50 \qquad b = 20$
$q = .50 \qquad n = 50$
$a = 30$

Substituting these values in the formula

$$X^2 = \frac{\left(a - \frac{p}{q} b \right)^2}{\frac{p}{q} n}$$

We get,

$$X^2 = \frac{\left(30 - \frac{.50}{.50}\ 20 \right)^2}{\frac{.50}{.50}\ 50} = 2$$

The number of degrees of freedom is 1. The value of X^2 for 1 degree of freedom at 5% level ofsignifiicance is 3.841. The calculated value is less than this figure. Hence the observed sex ratio during the given week is not significantly different from the general observation of sex ratio which is 1 : 1.

PROBLEM 17.21. *200 digits were chosen at random from a set of tables. The frequencies of the digits were :*

Digit	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Frequency	*18*	*19*	*23*	*21*	*16*	*25*	*22*	*20*	*21*	*15*

Use the X^2 test to assess the correctness of the hypothesis that the digits were distributed in equal numbers in the tables from which these were chosen.

SOLUTION. If we take the hypothesis that the digits were distributed in equal numbers in the tables as correct, the expected frequencies of the digits would be :

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	20	20	20	20	20	20	20	20	20	20

Substituting the observed and the expected frequencies in the formula

$$X^2 = S \left\{ \frac{(f - f_1)^2}{f_1} \right\}$$

We get,

$$X^2 = \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(16-20)^2}{20}$$
$$+ \frac{(25-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(21-20)^2}{20}$$
$$+ \frac{(15-20)^2}{20}$$
$$= \tfrac{1}{20} \{4+1+9+1+16+25+4+0+1+25\}$$
$$= 4.3.$$

The number of degrees of freedom
$= (c-1)(r-1) = (10-1)(2-1) = 9$

The 5% value of X^2 for 9 degrees of freedom is 16.919. The calculated value of X^2 which is 4.3 is thus much less than this figure. Hence the hypothesis seems reasonable and correct.

PROBLEM 17.22 *12 dice were thrown 4096 times and a throw of 6 was reckoned as success, the observed frequencies were as given below :—*

Success	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7 & over*
Frequencies	*447*	*1145*	*1181*	*796*	*380*	*115*	*24*	*8*

Find the value of X^2 on the hypothesis that the dice were unbiased and hence show that the data are consistent with this hypothesis so far as the X^2 test is concerned.

SOLUTION. If the hypothesis that the dice were unbiased is correct, the chance of getting a 6 with one die is $\frac{1}{6}$. Hence the chances with 12 dice of getting 12.6's 11,6's etc., are the successive terms in the binomial expansion $(\frac{1}{6}+\frac{5}{6})^{12}$. Hence the theoretical frequencies is 4096 throws are the terms in binomial expansion $(\frac{1}{6}+\frac{5}{6})^{12}\times 4096$ and are given below :—

Success	0	1	2	3	4	5	6	7 & over
Frequencies	459	1102	1212	808	364	116	27	8

Substituting the observed and the theoretical frequencies in the formula

$$X^2=S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get.

$$X^2=\frac{(447-459)^2}{459}+\frac{(1145-1102)^2}{1102}$$
$$+\frac{(1181-1212)^2}{1212}+\frac{(796-808)^2}{808}+\frac{(380-364)^2}{364}$$
$$+\frac{(115-116)^2}{116}+\frac{(24-27)^2}{27}+\frac{(8-8)^2}{8}$$
$$=\frac{144}{459}+\frac{1849}{2102}+\frac{961}{1212}+\frac{144}{808}+\frac{256}{364}+\frac{1}{116}+\frac{9}{27}+\frac{0}{8}=3.34$$

The number of degrees of freedom.

$=8-1=7.$

The value of X^2 for 7 degrees of freedom at 5% level of significance is 14.067. The calculated value is much less than it. This proves that the hypothesis is correct and the data are consistent with the hypothesis.

PROBLEM 17.23. *The number of males in each 106 eight-pig litters was found, and they are given by the following frequency distribution.—*

Number of males per litter	*0*	*1*	*2*	*3*	*4*	*5*	*6*	7	*8*	*Total*
Frequency	*0*	*5*	*9*	*22*	*25*	*26*	*14*	*4*	*1*	*106*

Assuming that the probability of animal being male or female is even (i.e., $p=q=\frac{1}{2}$), and the frequency distribution follows the binomial law, calculate the expected frequencies of the nine classes. Find also the value of X^2 to test the goodness of fit.

SOLUTION. Taking the hypothesis as correct, the chance of getting a male animal is $\frac{1}{2}$. Hence the chances of getting 8 males, 7 males, etc., in one eight-pig litter are the successive terms in the binomial expansion $(\frac{1}{2}+\frac{1}{2})^8$. Hence the theoretical frequencies in 106 eight-pig litters are the terms in $106\,(\frac{1}{2}+\frac{1}{2})^8$ which are as follows :—

Number of males per litter	0	1	2	3	4	5	6	7	8	Total
Frequency	.4	3.3	11.6	23.2	29	23.2	11.6	3.3	.4	106

To calculate X^2, we merge the first two and the last two cell frequencies into one so that they may not be less than 5. Now our data are

Number of males per litter	0 and 1	2	3	4	5	6	7 and 8
Observed Frequency	5	9	22	25	26	14	5
Theoretical Frequency	3.7	11.6	23.2	29	23.2	11.6	3.7

Substituting the observed and the theoretical frequencies in the formula

$$X^2=S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2=\frac{(5-3.7)^2}{3.7}+\frac{(9-11.6)^2}{11.6}+\frac{(22-23.2)^2}{23.2}+\frac{(25-29)^2}{29}$$
$$+\frac{(26-23.2)^2}{23.2}+\frac{(14-11.6)^2}{11.6}+\frac{(5-3.7)^2}{3.7}$$
$$=.46+.58+.06+.55+.34+.5+.46=2.95$$

The number of degrees of freedom

$=(c-)\,(r-1)=(7-1)\,(2-1)=6$

The table value of X^2 for 6 degrees of freedom at 5% level of significance is 12.592. The calculated value is much less than this figure. Hence we may say that there is correspondence between theory and practice or in other words, the fit is good.

Note:—One of the conditions for the application of the X^2 test is that no theoretical cell frequency should be small. It is hard to say what constitutes smallness, but 5 may be regarded as the minimum.

PROBLEM 17.24. *Ten coins are tossed 1024 times and the following frequencies observed.*

No. of heads	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	
Frequencies	*2*	*10*	*38*	*106*	*188*	*257*	*226*	*128*	*59*	*7*	*3*

How does this compare with a normal distribution ?

SOLUTION. Let us take as our hypothesis the supposition that the coin is unbiased and the distribution is normal. If this be true, the chance of getting heads etc., in a toss of ten coins are the successive terms in the binomial $(\frac{1}{2}+\frac{1}{2})^{10}$. So the theoretical frequencies in 1024 tosses are the terms in 1024 $(\frac{1}{2}+\frac{1}{2})^{10}$ and are as follows :—

No. of heads	0	1	2	3	4	5	6	7	8	9	10
Frequencies	1	10	45	120	210	252	210	120	45	10	1

To calculate X^2, we merge the first two and the last two cell frequencies in one so as to make them greater than 5.

No. of heads	0 and 1	2	3	4	5	6	7	8	9 and 10
Observed Frequencies	12	38	106	188	257	226	128	59	10
Theoretical Frequencies	11	45	120	210	252	210	120	45	11

Substituting the observed and the theoretical frequencies in the formula

$$X^2=S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2=\frac{(12-11)^2}{11}+\frac{(38-45)^2}{45}+\frac{(106-120)^2}{120}+\frac{(188-210)^2}{210}$$
$$+\frac{(257-252)^2}{252}+\frac{(226-210)^2}{210}+\frac{(128-120)^2}{120}$$
$$+\frac{(59-45)^2}{45}+\frac{(10-11)^2}{11}$$

$= .09+1.09+1.63+2.3+.1+1.22+.53+4.36+.09 = 11.41.$

The number of degrees of freedom

$= (c-1)(r-1) = (9-1)(2-1) = 8.$

The 5% value of X^2 for 8 degrees of freedom is 15.507. The calculated value is much less than this figure. Hence, the hypothesis appears reasonable and the observed distribution compares well with a normal distribution.

PROBLEM 17.25. *The following table shows the number of people interviewed by age groups and the number in each age group estimated to have peptic ulcers.*

Age group	*15—20*	*20—25*	*25—30*	*35—45*	*45—55*	*55—65*	*65—75*	*Total*
Nos. Interviewed	*199*	*300*	*1128*	*1375*	*1089*	*625*	*155*	*4871*
P. U. Cases	*1*	*8*	*38*	*96*	*105*	*56*	*12*	*316*

Do these figures justify the hypothesis that peptic ulcer is equally popular in all age groups ?

SOLUTION. If peptic ulcer was equally popular in all age groups then, in each $\left(\frac{316}{4871} \times 100\right)$ or 6.5% of the people should suffer from it. The observed and expected frequencies would be as follows :—

Age group	15—20	20—25	25—35	35—45	45—55	55—65	65—75
Observed Cases	1	8	38	96	105	56	12
Expected Cases	13	19.5	73	89	71	40.5	10

$$X^2 = \frac{(1-13)^2}{13} + \frac{(8-19.5)^2}{19.5} + \frac{(38-73)^2}{73} + \frac{(96-89)^2}{89}$$

$$+ \frac{(105-71)^2}{71} + \frac{(56-40.5)^2}{40.5} + \frac{(12-10)^2}{10} = 58.6$$

For 6 degrees of freedom the value of X^2 at 5% level of significance is 12.59. The calculated value is much higher than this and as such the results are significant and the hypothesis is not justified.

PROBLEM 17.26. *Two investigators draw samples from the same town in order to estimate the number of persons falling in the income groups "poorer," "middle class," "well-to-do" (The limits of the groups are defined in terms of money and are the same for both investigators.) Their results are as follows :—*

Investigator	*Income-group*			*Total*
	"Poorer"	*"Middle class"*	*"Well-to-do"*	
A	*140*	*100*	*15*	*255*
B	*140*	*50*	*20*	*210*
Totals	*280*	*150*	*35*	*465*

Show that the sampling technique of at least one of the investigators is suspect.

SOLUTION. Let us take as our hypothesis the supposition that the sampling technique of both the investigators is the same and beyond suspicion. If this be true, and as both the samples have been drawn from one population, the two samples should contain almost same proportion of "poorer," "middle class" and "well-to-do" people. On this basis the theoretical frequencies would be :—

Investigator	Income group			
	"Poorer"	"Middle class"	"Well-to-do"	Total
A	154	82	19	255
B	126	68	16	210
Total	280	150	35	465

Substituting the observed and the theoretical frequencies in the formula,

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(140-154)^2}{154} + \frac{(100-82)^2}{82} + \frac{(15-19)^2}{19} + \frac{(140-126)^2}{126}$$

$$+ \frac{(50-68)^2}{68} + \frac{(20-16)^2}{16}$$

$$= \frac{196}{154} + \frac{324}{82} + \frac{16}{19} + \frac{196}{126} + \frac{324}{68} + \frac{16}{16} = 13.38.$$

The number of degrees of freedom

$= (c-1)(r-1) = (3-1)(2-1) = 2.$

The value of X^2 for 2 degrees of freedom at 5% level of significance is 5.991 which is much less than the calculated value. Hence the hypothesis does not hold ground and suspicion arises in the sampling technique of at least one of the investigators.

PROBLEM 17.27. *In experiments on pea-breeding Mendel obtained the following frequencies of seeds : 315 round and yellow : 101 wrinkled and yellow ; 108 round and green ; 32 wrinkled and green. Total 556. Theory predicts that the frequencies should be in the proportions 9 : 3 : 3 : 1.*

Examine the correspondence between theory and experiment.

SOLUTION. On the theoretical considerations, the frequencies should be :

Round and yellow	...	313
Wrinkled and yellow	...	104
Round and green	...	104
Wrinkled and green	...	35

Substituting the observed and the theoretical frequencies in the formula,

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(315-313)^2}{313} + \frac{(101-104)^2}{104} + \frac{(108-104)^2}{104} + \frac{(32-35)^2}{35}$$

$$= \frac{4}{313} + \frac{9}{104} + \frac{16}{104} + \frac{9}{35} = .51.$$

The number of degrees of freedom is 4—1, or 3.

The value of X^2 for 3 degrees of freedom at 5 % level of significance is 7.81 The calculated value is much less than this figure. Hence there appears to be a very high degree of correspondence between theory and practice.

PROBLEM 17.28. *Genetic theory states that children having one parent of blood type M and the other of blood type N will always be of one of the three types M, MN, N and that the proportions of three types will on average be as 1 : 2 : 1 A report states that out of 300 children having one M parent and one N parent 30% were found to be type M, 45% type MN and remainder type N. Test the hypothesis by X^2 test.*

SOLUTION. If the genetic theory be true, 300 children having one M parent and one N parent, should be distributed as follows :—

Children of type M 75
Children of type MN 150
Children of type N 75

The observed frequencies are :

Children of type M 90
Children of type MN 135
Children of type N 75

Substituting the observed and the theoretical frequencies in the formula

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(90-75)^2}{75} + \frac{(135-150)^2}{150} + \frac{(75-75)^2}{75}$$
$$= \frac{225}{75} + \frac{225}{150} + \frac{0}{75} = 4.5.$$

The number of degrees of freedom is 3—1, or 2. The 5% value of X^2 for 2 degrees of freedom is 5.991. The calculated value is less than this figure Hence there can be said to be correspondence between theory and practice and the hypothesis holds ground so far as the X^2 test is concerned.

PROBLEM 17.29. *Given the following actual and theoretical normal frequencies (total 400) test the goodness of fit (degree of freedom 10).*

Actual 4, 11, 17, 29, 43, 56, 58, 63, 61, 25, 20, 9, 4

Theoretical 4.6, 7.9, 16.8, 30.3, 44.7, 59.1, 65 60.4, 47.5, 31.16, 18.25, 8.8, 5.3

SOLUTION. Substituting the actual and the theoretical frequencies in the formula,

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(4-4.6)^2}{4.6} + \frac{(11-7.9)^2}{7.9} + \frac{(17-16.8)^2}{16.8} + \frac{(29-30.3)^2}{30.3}$$
$$+ \frac{(43-44.7)^2}{44.7} + \frac{(56-59.1)^2}{59.1} + \frac{(58-65)^2}{65} + \frac{(63-60.4)^2}{60.4}$$
$$+ \frac{(61-47.5)^2}{47.5} + \frac{(25-31.16)^2}{31.16} + \frac{(20-18.25)^2}{18.25}$$
$$+ \frac{(9-8.8)^2}{8.8} + \frac{(4-5.3)^2}{5.3}$$

$$= \frac{.36}{4.6} + \frac{9.61}{7.9} + \frac{.04}{16.8} + \frac{1.69}{30.3} + \frac{2.89}{44.7} + \frac{9.61}{59.1}$$
$$+ \frac{49}{65} + \frac{6.76}{60.4} + \frac{182.25}{47.5} + \frac{37.95}{31.16} + \frac{3.06}{18.25} + \frac{.04}{8.8} + \frac{1.69}{5.3} = 7.6821$$

The value of X^2 for 10 degrees of freedom at 5% level of significance is 18.307. The calculated value is much less than this figure. Hence the data denote a reasonably good fit.

PROBLEM 17.30. *Test the following data for goodness of fit (volume of trading on New Delhi Stock Exchange expressed as percentage of straight line trend, 1957—1973).*

Class Marks	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Observed frequencies	*11*	*35*	*50*	*48*	*24*	*15*	*9*	*7*	*3*	*1*	*1*
Theoretical frequencies	*15*	*29*	*48*	*43*	*35*	*21*	*9*	*3*	*1*	*0*	*0*

Here $n=204$. Is the theoretical frequency curve a good fit to the observed data ?

SOLUTION. Substituting the observed and the theoretical frequencies to the formula

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(11-15)^2}{15} + \frac{(35-29)^2}{29} + \frac{(50-48)^2}{48} + \frac{(48-43)^2}{43} + \frac{(24-35)^2}{35}$$
$$+ \frac{(15-21)^2}{21} + \frac{(9-9)^2}{9} + \frac{(7-3)^2}{3} + \frac{(3-1)^2}{1} + \frac{(1-0)^2}{0} + \frac{(1-0)^2}{0}$$
$$= \frac{16}{15} + \frac{36}{29} + \frac{4}{48} + \frac{25}{43} + \frac{121}{35} + \frac{36}{21} + \frac{0}{9} + \frac{16}{3} + 4 + 1 + 1 = 19.5$$

The number of degrees of freedom is 11—1, or 10. The value of X^2 for 10 degrees of freedom at 5% level of significance is 18.307 and at 1% level of significance is 23.21. Therefore for the calculated value, p lies between .01 to .05. Hence, the conclusion is that the observed data present a poor fit to the theoretical frequency curve.

PROBLEM 17.31. *In a certain cross the types represented by BC, Bc, bC and bc are expected to occur in a 9 : 3 : 3 : 1 ratio. The actual frequencies obtained were :*

BC	*Bc*	*bC*	*bc*
102	*16*	*35*	*7*

SOLUTION. According to a theoretical 9 : 3 : 3 : 1 ratio, the theoretical frequencies would be :

BC	Bc	bC	bC
90	30	30	10

Substituting the observed or actual frequencies and the theoretical frequencies in the formula.

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(102-90)^2}{90} + \frac{(16-30)^2}{30} + \frac{(35-30)^2}{30} + \frac{(7-10)^2}{10}$$
$$= \frac{144}{90} + \frac{196}{30} + \frac{25}{30} + \frac{9}{10} = 9.87.$$

The number of degrees of freedom is 4—1, or 3. The value of X^2 for 3 degrees of freedom is 7.82 when $p=0.05$; and 11.34 when $p=0.01$. Therefore the calculated value of X^2 lies between these two values, hence the fit is poor.

PROBLEM 17.32. *Test the goodness of fit of observation to theory for the following ratios :*

	Observed	*Values*	*Theoretical*	*Values*
	A	*a*	*A*	*a*
(*1*)	*134*	*34*	*3*	*1*
(*2*)	*240*	*120*	*3*	*1*
(*3*)	*76*	*56*	*1*	*1*
(*4*)	*243*	*13*	*15*	*1*

SOLUTION. According to the given ratios, theoretical frequencies for the four observations would be :

	A	*a*
(1)	126	42
(2)	270	90
(3)	66	66
(4)	240	16

Substituting the observed and the theoretical frequencies to the formula

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

Where f represents the actual or observed frequencies and f_1 the expected or theoretical frequencies

We get,

$$X^2 \text{ for the first observation} = \frac{(134-126)^2}{126} + \frac{(34-42)^2}{42}$$

$$= \frac{64}{126} + \frac{64}{42} = 2.$$

$$X^2 \text{ for the 2nd observation} = \frac{(240-270)^2}{270} = \frac{(120-90)^2}{90}$$

$$= \frac{900}{270} + \frac{900}{90} = 13.33.$$

$$X^2 \text{ for the third observation} = \frac{(76-66)}{66} + \frac{(56-66)^2}{66}$$

$$= \frac{100}{66} + \frac{100}{66} = 3.03.$$

$$X^2 \text{ for the fourth observation} = \frac{(243-240)^2}{240} + \frac{(13-16)^2}{16}$$

$$= \frac{9}{240} + \frac{9}{16} = .6$$

The number of degrees of freedom in each case is 1. For 1 degree of freedom the value of X^2 at 5% level of significance is 3.841. Judging from this value, the fit in the fourth observation is excellent, in the first and the third good, while in the second, it is extremely poor.

PROBLEM 17.33. *Number of previous children by present husband.*

0	*1*	*2*	*3*	*4*	*Total*
17167	*1861*	*124*	*2*	*1*	*19155*

Compute the expected frequencies on the assumption that this is a Poisson distribution and apply the X^2 test of goodness of fit to the data.

SOLUTION. The Poisson distribution is given by

$$P(x) = \frac{mx}{x!}\mathrm{e}^{m} \text{ where } m \text{ is the mean.}$$

The mean of the above distribution $= m$

$$\frac{(0\times 17167)+(1\times 1861)+(2\times 124)+(3\times 2)+(4\times 1)}{19{,}155}$$

$$=\frac{0+1861+248+6}{19{,}155}=\frac{2{,}119}{19{,}155}=0.1106$$

$\therefore$ The Poisson distribution is given by $\frac{(0.1106)}{x\,!}e^{-0\cdot 1186}$

Calculation of the theoretical frequencies :

$$e^{-0\cdot 1108}=\left(1-0.1106+\frac{(0.1106)^2}{2\,!}-\frac{(0.1106)^3}{3\,!}+\frac{(0.1106)^4}{4\,!}\cdots\cdots\right)$$

$$=1-0.1106+0.0061-0.0002=1.0061-0.1108=0.8953$$

$P(_0)=\frac{(.1106)}{0\,!}e^{-1103}=0.8953$ $\qquad\therefore f_1\,(^0)=0.8953\times 19155=17150$

$P(_1)=\frac{(.1106)}{1\,!}e^{-\cdot 1106}=0.8953\times 0.1106=0.0990$ $\therefore f_1\,(^1)=0.0990\times 19155=1896$

$P(_2)=\frac{(.1106)^2}{2\,!}e^{-\cdot 1106}=0.99\times .0553$ $\qquad\therefore f_1\,(^2)=0.0055\times 19155=105=0.0055$

$=0.0055\cdot$

$P(_3)=\frac{(.1106)^3}{3\,!}e^{-\cdot 1166}=.0055\times .0369$ $\quad\therefore f_1\,(^3)=0.0002\times 19155=4$

$=0.0002$

$P(_4)=\frac{(.1106)^4}{4\,!}e^{-\cdot 1106}=0.0002\times .0276$ $\therefore f_1\,(^4)=0.000055\times 19155=0$

$0=0.000055$

The theoretical frequencies are 17150, 1896, 105, 4 and 0.

Since the frequencies in each of the two cells are less than 10, we group the last three cells in the calculation of X^2.

f	:	17167	1861	127
f_1	:	17150	1896	109
$f-f_1$	:	17	—35	18
$(f-f_1)^2$	:	289	1225	324
$\frac{(f-f_1)^2}{f_1}$	:	0.0168	0.6461	2.9725

$$\therefore X^2=\Sigma\,\frac{(f-f_1)^2}{f_1}=3.6354=3.64 \text{ (approx.)}$$

No. of classes $=3$

$\therefore$Degrees of freedom $=3-1=2$

and we estimate one perimeter from the data.

For 1 degree of freedom, .05 level of value of $X^2=3.84$.

The observed value of X^2 is less than this and, therefore, not significant.

PROBLEM 17.34. *192 families (for each of which the possibility of an albino child being born is otherwise established) had the following distribution of albinos among the first children.*

No. of albinos	*No. of families*
0	*77*
1	*90*
2	*20*
3	*5*
	192

Find the expected frequencies on the basis of a theoretical probability 0.25 of a child being born an albino and test the goodness of fit. Estimate the probability of a child being born an albino from the given data and test whether it significantly deviates from the hypothetical value 0.25.

SOLUTION. The probability of 0, 1, 2 and 3 albinos are given by terms of expansion of $(\frac{3}{4}+\frac{1}{4})^3$

They are $0—\frac{27}{64}$
$1—\frac{27}{64}$
$2—\frac{9}{64}$
$3—\frac{1}{64}$

$\therefore$ Theoretical frequencies are $192\,(\frac{27}{64}, \frac{27}{64}, \frac{9}{64}, \frac{1}{64})$ or 81, 81. 27 and 3.
Testing Goodness of fit.

Since in the last group, the frequencies are less than 10, we combine 3rd 4th groups.

f	f_1	$f-f_1$	$(f-f_1)^2$	$\frac{(f-f_1)^2}{f_1}$
77	81	—4	16	0.19753
90	81	9	81	1.00000
25	30	—5	25	0.83333
Total 192	192	0		$2.03053 = X^2$

No. of classes = 3

$\therefore$ Degrees of freedom = 3—1 = 2

For 2 degrees of freedom .05 level value of X^2 is 5.99. Since the observ value of X^2 is less than this value and therefore not significant.

We can use the X^2 test for testing the hypothesis that a number of differe samples correspond to variables having the same distribution without specify the distribution.

PROBLEM 17.15 *The following table gives the number of new born bo and girls far each of twelve months. Are these data consistent with the hypothe that there is during all twelve months a consistent probability that new born ch will be a boy ?*

Months	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*To*
Boys	*3743*	*3550*	*4017*	*4173*	*4117*	*3944*	*3964*	*3797*	*3712*	*3512*	*3392*	*3761*	*456*
Girls	*3537*	*3407*	*3866*	*3711*	*3775*	*3665*	*3621*	*3596*	*3491*	*3391*	*3160*	*3371*	*4259*
Total	*7280*	*6957*	*7883*	*7884*	*7892*	*7609*	*7585*	*7393*	*7203*	*6903*	*6552*	*7132*	*8872*

SOLUTION. Our hypothesis is that there is during all twelve months constant probability that a new born child will be a boy.

On this hypothesis the estimated value of this probability

$$= \frac{45682}{88273} = 0.5173$$

For calculating X^2 we note that each of the 12 observations gives two groups (1) cases where a new born child is a boy and (2) cases where a new born child is not a boy. The estimated probability of the latter event = 1—0.5175 = 0.4825.

The theoretical frequencies for each month are given in the following table :

Calculation of X^2

$$\text{Contribution of one item to } X^2 = \frac{[(f_i - n_i p)^2}{n_i p} + \frac{(n_i - f_i - n_i)(1-p)]^2}{n_i(1-p)}$$

$$= \frac{(f_1 - pn_i)^2}{p(1-p)\,x_i}$$

$$\therefore X^2 = \frac{1}{p(1-p)} \Sigma \frac{(f_i - pn_i)^2}{n_i}$$

Months	Total (n_i)	Boys (f_i)	Theoretical Frequency (Boys) pn_i	$(f-pn_i)^2$	$\frac{(f_i-pn_i)^2}{n_i}$
1	7280	3743	3767	$24^2=576$	0.0792
2	6957	3550	3600	$50^2=2500$	0.3593
3	7883	4017	4079	$62^2=3844$	0.4876
4	7884	4173	4080	$93^2=8649$	1.0972
5	7892	4117	4084	$33^2=1089$	0.2645
6	7609	3944	3938	$6^2=36$	0.0047
7	7585	3964	3925	$39^2=1521$	0.2005
8	7393	3797	3826	$29^2=841$	0.1137
9	7203	3712	3728	$16^2=256$	0.0355
10	7903	3512	3572	$60^2=3600$	0.5021
11	6552	3392	3391	$1^2=1$	0.0002
12	7132	3761	3691	$70^2=4900$	0.6871
					3.8316

$$\therefore X^2 = \frac{1}{0.5175 \times 0.4825} \times 3.8316 = 15.333.$$

For each month we have two groups and, therefore, 2—1=1 degree of freedom. There are 12 months and one parameter (probability has been estimated).

Degrees of freedom=12—1=11.

For 11 degrees of freedom, .05 level value of $X^2=19.68$. The observed value X^2 is less than this and, therefore, it is not significant. The agreement with the hypothesis of a constant probability is quite satisfactory.

PROBLEM 17.36. *From the adult male population of seven large cities, random samples of the sizes indicated below were taken and the numbers of married and single men recorded. Do the data indicate any significant variation among the cities in the tendency of men to marry ?*

City	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*Total*
Married	*133*	*164*	*155*	*106*	*153*	*123*	*146*	*980*
Single	*36*	*57*	*40*	*37*	*55*	*33*	*36*	*294*
Total	*169*	*221*	*195*	*143*	*208*	*156*	*182*	*1274*

SOLUTION. Our hypothesis is that there is no significant variation among the cities in the tendency of men to marry.

On this hypothesis the ratio between the married and single men in all the 7 cities will be the same and the estimated value of the ratio will be equal to the ratio in the total.

The estimated value of ratio=980 ; 294=10 : 3

Cities	Total (n_i)	Married (f_i)	Theoretical Frequency Married (np_i)	$(f_1-np_i)^2$	$\frac{(f_i-np_i)^2}{n_i}$
A	169	133	130	$3^2=9$	0.053
B	221	164	170	$6^2=36$	0.163
C	195	155	150	$5^2=25$	0.128
D	143	106	110	$4^2=16$	0.112
E	208	153	160	$7^2=49$	0.235
F	156	123	120	$3^2=9$	0.058
G	182	146	140	$6^2=36$	0.198
					0.947

$$\therefore X^2 = \frac{1}{p(1-p)} \Sigma \frac{(f-pn_i)^2}{n_i}$$

$$=\frac{1}{\frac{3\times 10}{13\times 13}}\times 0.947=\frac{169}{30}\times 0.947=\frac{16.0043}{3}=5.335$$

For each city we have two groups and, therefore, 2—1=1 degree of freedom. One parameter has been estimated, therefore, degrees of freedom = 7—1=6.

For 6 degrees of freedom. .05 level value of $X^2=12.59$. The observed value X^2 is less than this and, therefore, the test does not offer any evidence against the hypothesis.

PROBLEM 17.37. *A certain hypothesis was tested by two similar experiments, which gave $X^2=14.7$ for 9 degrees of freedom and $X^2=14.9$ for 11 degrees of freedom. Show that the two experiments combined give less reason for confidence in the hypothesis than either experiment alone.*

SOLUTION. X has got additive property.

Thus the value of X^2 for two experiments combined is $14.7+14.9=29.6$ with $9+11=20$ degrees of freedom.

Now from tables we find :

For 9 degrees of freedom $X^2=14.69$ is significant at .10 level

For 11 degrees of freedom $X=14.63$ is significant at .20 and $X^2=17.28$ is significant at .10 level,

and for 20 degrees of freedom, $X^2=28.41$ is significant at .10 level .05 level values of X^2 for 9, 11 and 20 degrees of freedom are 16.92, 19.68 and 31.41 respectively. None of the observed values of X^2 is significant at .05 level, all observed values being less than the theoretical values. But when we consider these at .10 level of significance we find that the observed values when compared with theoretical values, are not significant, but the combined value is significant at this level. Thus two experiments give less reason for confidence in the hypothesis than either experiment alone.

PROBLEM 17.38. *250 samples of coal, distributed according to ash percentage.*

Percent Ash	*9.00— 9.99*	*10.00—10.99*	*11.00—11.99*	*12.00—12.99*	*13.00—13.99*
Frequency	*1*	*3*	*3*	*9*	*13*
Percent Ash	*14.00—14.99*	*15.00—15.99*	*16.00—16.99*	*17.00—17.99*	*18.00—18.99*
Frequency	*27*	*28*	*39*	*42*	*34*
Percent Ash	*19.00—19.99*	*20.00—20.99*	*21.00—21.99*	*22.00—22.99*	*23.00—23.99*
Frequency	*19*	*14*	*10*	*4*	*3*
Percent Ash	*24.00—24.99*	*25.00—25.99*	*Total*		
Frequency	—	*1*	*250*		

Can this be regarded as a sample from normal population ?

SOLUTION. (I) Calculation of the mean and standard deviation :

Mid value	d	f	df	d^2f
9.495	—8	1	— 8	64
10.495	—7	3	— 21	147
11.495	—6	3	— 18	108
12.495	—5	9	— 45	225
13.495	—4	13	— 52	208
14.495	—3	27	— 81	243
15.495	—2	28	— 56	112
16.495	—1	39	— 39	39
17.495	0	42	—320	0
18.495	1	34	34	34
19.495	2	19	38	76
20.495	3	14	42	126
21.495	4	10	40	160
22.495	5	4	20	100
23.495	6	3	18	108
24.495	7	0	0	0
25.495	8	1	8	64
			200	
			—120	1814

$$x=17.495-\tfrac{120}{250}$$
$$=17.495-0.480$$
$$=17.015$$
$$\sigma^2=\tfrac{1814}{250}-(.48)^2$$
$$=7.2560-0.2304$$
$$=7.0256$$

Corrected σ^2

$$=72.256-.0816$$

$$=6.944$$
$$\therefore \sigma=2.635$$

(*ii*) Next, we have to find out the frequencies of the normal curve with the same mean, standard deviation and total frequency.

In this connection we note that any of the class frequencies should not be less than 10, and therefore, we pool the first four and the last five classes :

	Standardized	Theoretical			
Class limits	class limits $\left(\frac{x-x}{\sigma}\right)$	Frequency		$f-f_1$	$\frac{(f-f_1)^2}{f_1}$
		f	f_1		
<13.00—	<—1.524	16	16.012	—0.012	0.000
13.00—	—1.524	13	15.640	—2.640	0.446
14.00—	—1.144	27	24.023	—2.977	0.369
15.00—	—0.765	28	31.878	—3.878	0.472
16.00—	—0.385	39	36.850	2,150	0.125
17.00—	—0.006	42	37.015	4.985	0.671
18.00—	0.374	34	32.080	1.920	0.115
19.00—	0.753	19	24.285	—2.285	1.150
20.00—21.00	1.133—1.512	14	15.875	—1.875	0.222
21.00	1.512	18	16.342	1.658	0.168
		250	250.000	0.000	3.738

$$\therefore X^2=\Sigma\frac{(f-f_1)^2}{f_1}=3.738$$

There are ten classes; and two parameters (x and σ) were estimated from the data.

$\therefore$ Degrees of freedom $=10-1-2=7$.

For 7 degrees of freedom, 0 05 level value of X^2 is 14.067

The observed value of X^2 is much less than this and, therefore. data agree very well with the hypothesis.

X² AND NORMAL DISTRIBUTION

PROBLEM 17.39. *In a certain 'thought transference' experiment a pack of cards with 5 different suits is used. The subject has to name the suit of a card picked at random and shown to the thought transferor in another room. On one occasion 228 correct results were obtained in a thousand such trials. Does this indicate that the subject was not merely guessing ?*

SOLUTION. This problem can be solved by two methods. We shall solve it by these and indicate an important property of x^2 distribution.

(*i*) *By applying the normal distribution :*

We find that the mean value $= \frac{1}{5} \times 1000 = 200$ and the standard deviation $= 12.65$.

Now the number of correct results is 228. Since the normal distribution is a continuous distribution, 228 should be taken to cover the interval 227.5—228.5.

The standardized variate $= \dfrac{x-\bar{x}}{\sigma} = \dfrac{227.5-200}{12.65} = \dfrac{27.5}{12.65} = 2.174$

The probability that the deviation is more than $2.174 = 0.5 - 0.4850 = 0,0150$

The probability is small, and it is very unlikely that he was guessing.

(*ii*) *By applying the X^2 test :*

	f	f_1	$f-f_1$	$(f-f_1)^2$	$\dfrac{(f-f_1)^2}{f_1}$
Right results	228	200	28	784	3.92
Wrong results	772	800	—28	784	0.98
	1000	1000	0		$4.90 = X^2$

Degrees of freedom = 1

For 1 degree of freedom, .05 level value of $X^2 = 3.84$. The difference is significant and it is unlikely that he was guessing.

Now we come to the next point. When there is only one degree of freedom, $\sqrt{X^2}$ is a normal deviate with zero as the mean and one as the standard deviation.

Thus $X^2 = \sqrt{4.90.}$

$\sqrt{X^2} = \sqrt{4.90} = 2.214$

And the probability that the deviation will exceed by this amount can be found from the table of normal distribution. The value is $0.5 - 0.4865 = 0.0135$.

The discrepancy between this result and the result obtained in (*i*) is accounted by the fact x^2 is a continuous distribution and, therefore, 28 must be taken as the interval 27.5—28.5. This correction—called Yate's Correction—must be made whenever the degree of freedom is one.

Applying this correction we find :

	f	f_1	$f-f_1$	$(f-f_1)^2$	$\dfrac{(f-f_1)^2}{f_1}$
Right	227.5	200	27.5	756.25	3.7812
Wrong	772.5	800	—25.5	756.25	0.9453
	1000.0	1000	0		$4.7265 = X^2$

$$\therefore \sqrt{X^2} = 4.7265 = 2\text{-}17 = \frac{x-x}{\sigma} = \frac{x-0}{1} = x$$

The probability that the deviation will exceed by this amount can be found out from the table of Normal Distribution. This value is $0.5 - 0.485 = 0.015$ which agrees with the value arrived at in (*i*).

PROBLEM 17.40. *In an experiment, it was found that for 36 degrees of freedom, the observed value of X^2 was 60. Find out value of P, i.e., the probability of exceeding this value in random sampling.*

SOLUTION. It will be seen that the problem cannot be solved by looking he tables because they tabulate values of P upto 30 degrees of freedom

(*i*) To solve this problem we use a property of X^2 *i.e.* when degrees of free- >30, $\sqrt{2X^2}$ is distributed approximately normally—its mean being $\sqrt{2\nu-1}$ standard deviation$=1$.

Now $X^2=60 \therefore \sqrt{2X^2}=\sqrt{120}=10.954$

The mean of the distribution of $\sqrt{2X^2}\ \sqrt{2\nu-}1=\sqrt{72-1}$

$$=\sqrt{71}=8.426$$

And standard deviation$=1$

$\therefore$Standard deviation $=\dfrac{x-\overline{x}}{\sigma}=\dfrac{10.954-8.426}{1}=2.528$

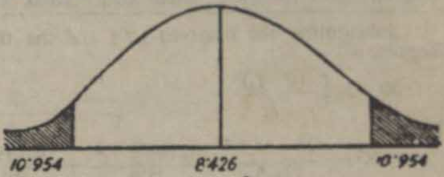
In the table of normal distribution we find that the probability that there ll be a deviation greater than

$528=2\ (0.5-0.49426)$

$=0.00578\times 2$

$=0.01148$

haded area in the figure)



(*ii*) There is another rmula which is more accurate.

$3\sqrt{\dfrac{X^2}{\nu}}$ is normally distributed about mean $1-\dfrac{2}{9\nu}$ and with standard viation $\sqrt{\dfrac{2}{9\nu}}$

Now $X^2=60$ and $\nu=36$

$\therefore 3\sqrt{\dfrac{60}{36}}$ is the value of a normal variate, the mean of the distribution is $-\dfrac{2}{9\times 36}$ and variance is $\dfrac{2}{9\times 36}$

$x=\dfrac{1.710}{1.442}=1.1858;\ \overline{x}=1-0.00617=0.99383$

d $\sigma=\frac{1}{18}\sqrt{2}=0.05556\times 1.414=0.07856$

Standardized Deviation$=\dfrac{1.1858-0.9938}{0.0786}=\dfrac{0.1920}{0.0786}=2.44$

In the table of normal distribution we find that the probability that the viation will be greater than $2.44=2\ (0.5-0.4927)$

$$=2\times 0.0073=P=0.0146.$$

CO-EFFICIENT OF CONTINGENCY

PROBLEM 17.41. *The following table shows the association among 1000 hool boys between their general ability and their mathematical ability. Calculate e co-efficient of contingency between the two.*

Mathematical ability	*General ability*		
	Good	*Fair*	*Poor*
Good	44	22	4
Fair	265	257	178
Poor	41	91	98

SOLUTION. Proceeding on the hypothesis that the two attributes, General ability and the Mathematical ability are independent, we get the followi table of theoretical frequecies :

Maths. ability	General ability			
	Good	Fair	Poor	Total
Good	24.5	25.9	19.6	70
Fair	245.0	259.0	196.0	700
Poor	80.5	85.1	64.4	230
Total	350.0	370.0	280.0	1000

Substituting the observed (f) and the theoretical frequencies (f) in formula.

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(44-24.5)^2}{24.5} + \frac{(22-25.9)^2}{25.9} + \frac{(4-19.6)^2}{19.6}$$
$$+\frac{(265-245)^2}{245} + \frac{(257-259)^2}{259} + \frac{(178-196)^2}{196}$$
$$+\frac{(41-80.5)^2}{80.5} + \frac{(91-85.1)^2}{85.1} + \frac{(98-64.4)^2}{64.4}$$
$$= \frac{380.25}{24.5} + \frac{15.21}{25.9} + \frac{243.36}{19.6} + \frac{400}{245} + \frac{4}{259}$$
$$+\frac{324}{196} + \frac{1560.25}{80.5} + \frac{34.81}{85.1} + \frac{1128.96}{64.4}$$
$$= 15.52 + .59 + 12.42 + 1.63 + .02 + 1.65 + 19.38 + .409 + 17.53$$
$$= 69.14$$

Substituting the values of X^2 and N in the Pearson's formula

$$C = \sqrt{\frac{X^2}{N+X^2}}$$

Where C represents the co-efficient of Mean Square Contingency

We get,

$$C = \sqrt{\frac{69.15}{1000+69.4}} = .254$$

Thus, the co-efficient of contingency is .254.

PROBLEM 17.42. *Given the following contingency table for Hair colo und Eye colour, find the value of C. Is there good association between the two ?*

Eye colour	*Hair colour*			*Total*
	Fair	*Brown*	*Black*	
Blue	*15*	*5*	*20*	*40*
Grey	*20*	*10*	*20*	*50*
Brown	*25*	*15*	*20*	*60*
Total	*60*	*30*	*60*	*150*

SOLUTION. Proceeding on the hypothesis that the two attributes, Eye colour and the Hair colour given in the contingency table are not associated, we get the following theoretical frequencies.

Eye colour	Hair colour			Total
	Fair	Brown	Black	
Blue	16	8	16	40
Grey	20	10	20	50
Brown	24	12	24	60
Total	60	30	60	150

Substituting the observed and the theoretical frequencies in the formula

$$Xr = S\left\{\frac{(f-_1)^2}{f_1}\right\};$$

We get,

$$X^2 = \frac{(15-16)^2}{16} + \frac{(5-8)^2}{8} + \frac{(20-16)^2}{16}$$

$$+ \frac{(20-20)^2}{20} + \frac{(10-10)^2}{10} + \frac{(20-20)^2}{20}$$

$$+ \frac{(25-24)^2}{24} + \frac{(15-12)^2}{12} + \frac{(20-24)^2}{24}$$

$$= \frac{1}{16} + \frac{9}{8} + \frac{16}{16} + \frac{0}{20} + \frac{0}{10} + \frac{0}{20} + \frac{1}{24} + \frac{9}{12} + \frac{16}{24} = 3.7$$

Substituting the value of X^2 and N in the Pearson's formula

$$C = \sqrt{\frac{X^2}{N+X}};$$

where C represents the co-efficient of Mean Square Contingency.

We get,

$$C = \sqrt{\frac{3.7}{150+3.7}} = .15.$$

Thus, the co-efficient of Mean Square Contingency is .15 which indicates that the association between Eye colour and Hair colour is poor.

PROBLEM 17.43. *The following table shows the association among 1000 criminals between their weight and mentality. Calculate the co-efficient of contingency between the two :*

Mentality	*Weight in lbs.*					*Totals*
	90—120	*120—130*	*130—140*	*140—150*	*150 upward*	
Normal	*50*	*102*	*198*	*210*	*240*	*800*
Weak	*30*	*38*	*72*	*30*	*30*	*200*
Totals	*80*	*140*	*270*	*240*	*270*	*1000*

SOLUTION. Assuming that the two attributes weight and mentality are independent, we get the following theoretical cell frequencies.

Mentality	Weight in lbs.					Totals
	90—120	120—130	130—140	140—150	150 upwards	
Normal	64	112	216	192	216	800
Weak	16	28	54	48	54	200
Totals	80	140	270	240	270	1000

Substituting the observed and the theoretical frequencies in the formula.

$$X^2=S\left\{\frac{(f-f_1)^2}{f_1}\right\};$$

We get,

$$X^2=\frac{(50-64)^3}{64}+\frac{(102-112)^2}{112}+\frac{(198-216)^2}{216}+\frac{(210-192)^2}{192}$$
$$+\frac{(240-216)^2}{216}+\frac{(30-16)^2}{16}+\frac{(38-28)^2}{28}+\frac{(72-54)^2}{54}$$
$$+\frac{(30-48)^2}{48}+\frac{(30-54)^2}{54}$$
$$=\frac{196}{64}+\frac{100}{112}+\frac{324}{216}+\frac{324}{192}+\frac{576}{216}+\frac{196}{16}+\frac{100}{28}$$
$$+\frac{324}{54}+\frac{324}{48}+\frac{576}{54}=49.047.$$

Substituting the value of X^2 and N in the Pearson's formula

$$C=\sqrt{\frac{X^2}{N+X^2}}$$

Where C represents the co-efficient of Mean Square Contingency

We get,

$$C=\sqrt{\frac{49.047}{1000+49.047}}=.216$$

Thus, the co-efficient of contingency between weight and mentality of 1000 criminals is .216.

PROBLEM 17.44. *The data in the following table were obtained in a cross between a rust resistant and a susceptible variety of oats. The F_3 families were compared for reaction to rust in the seedling stage, and in the field under ordinary epidemic conditions.*

Table

Classification of Seedling and Field Reactions of 900 F_3 Families of oats.

Field Reaction	*Seedling Reaction*		
	Resistant	*Segregating*	*Susceptible*
Resistant	*112*	*51*	*37*
Segregating	*47*	*404*	*49*
Susceptible	*12*	*13*	*175*

Test the significance of the association in this table, and calculate the co-efficient of contingency.

SOLUTION. Assuming that the two attributes under consideration are ndependent we get the following theoretical cell frequencies.

Field	Seedling Reaction			Totals
	Resistant	Segregating	Susceptible	
Resistant	38	104	58	200
Segregating	95	260	145	500
Susceptible	38	104	58	200
Total	171	468	261	900

Substituting the observed and the theoretical frequencies in the formula

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$\begin{aligned} X^2 &= \frac{(112-38)^2}{38} + \frac{(51-104)^2}{104} + \frac{(37-58)^2}{58} \\ &+ \frac{(47-95)^2}{95} + \frac{(404-260)^2}{260} + \frac{(49-145)^2}{145} \\ &+ \frac{(12-38)^2}{38} + \frac{(13-104)^2}{104} + \frac{(175-58)^2}{58} \\ &= \frac{5476}{38} + \frac{2809}{104} + \frac{441}{58} + \frac{2304}{95} + \frac{20736}{260} \\ &+ \frac{9216}{145} + \frac{676}{38} + \frac{8281}{104} + \frac{13689}{58} = 679{\cdot}72. \end{aligned}$$

Substituting the value of X^2 and N in the Pearson's formula.

$$C = \sqrt{\frac{X^2}{N + X^2}}$$

Where C represents the co-efficient of Mean Square Contingency.

We get,

$$C = \sqrt{\frac{679{\cdot}72}{900 + 679{\cdot}72}} = {\cdot}656$$

Thus, the co-efficient of contingency is ·656 which is indicative of the fact that the association is significant.

PROBLEM 17·45. *1000 subjects of English, French, German, Italian and Spanish nationality were asked to name their preference among the music of those five nationalities. The results were as follows—*

Nationality of subject	*Nationality of music preferred*					*Total*
	English	*French*	*German*	*Italian*	*Spanish*	
English	*32*	*16*	*75*	*47*	*30*	*200*
French	*10*	*67*	*42*	*41*	*40*	*200*
German	*12*	*23*	*107*	*36*	*22*	*200*
Italian	*16*	*20*	*44*	*76*	*44*	*200*
Spanish	*8*	*53*	*30*	*43*	*66*	*200*
Total	*78*	*179*	*298*	*243*	*202*	*1000*

Discuss the association between the nationality of the subject and the nationality of the music preferred.

SOLUTION. Let us take as our hypothesis the supposition that the two attributes, *viz.*, nationality of music preferred and nationality of subject are independent. If this be true, the theoretical cell frequencies would be :

Nationality of subject	Nationality of music preferred					Totals
	English	French	German	Italian	Spanish	
English	15·6	35·8	59·6	48·6	40·4	200
French	15·6	35·8	59·6	48·6	40·4	200
German	15·6	35·8	59·6	48·6	40·4	200
Italian	15·6	35·8	59·6	48·6	40·4	200
Spanish	15·6	35·8	59.6	48·6	40·4	200
Totals	78	179	298	243	202	1000

Substituting the observed and the theoretical frequencies in the formula,

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$\begin{aligned}
X^2 = {} & \frac{(32-15{\cdot}6)^2}{15{\cdot}6} + \frac{(10-15{\cdot}6)^2}{15{\cdot}6} + \frac{(12-15{\cdot}6)^2}{15{\cdot}6} + \frac{(16-15{\cdot}6)^2}{15{\cdot}6} + \frac{(8-15{\cdot}6)^2}{15{\cdot}6} \\
& + \frac{(16-35{\cdot}8)^2}{35{\cdot}8} + \frac{(67-35{\cdot}8)^2}{35{\cdot}8} + \frac{(23-35{\cdot}8)^2}{35{\cdot}8} + \frac{(20-35{\cdot}8)^2}{35{\cdot}8} + \frac{(53-35{\cdot}8)^2}{35{\cdot}8} \\
& + \frac{(75-59{\cdot}6)^2}{59{\cdot}6} + \frac{(42-59{\cdot}6)^2}{59{\cdot}6} + \frac{(107-59{\cdot}6)^2}{59{\cdot}6} + \frac{(44-59{\cdot}6)^2}{59{\cdot}6} \\
& + \frac{(30-59{\cdot}6)^2}{59{\cdot}6} + \frac{(47-48{\cdot}6)^2}{48{\cdot}6} + \frac{(41-48{\cdot}6)^2}{48{\cdot}6} + \frac{(36-48{\cdot}6)^2}{48{\cdot}6} \\
& + \frac{(76-48{\cdot}6)^2}{48{\cdot}6} + \frac{(43-48{\cdot}6)^2}{48{\cdot}6} \\
& + \frac{(30-40.4)^2}{40.4} + \frac{(40-40.4)^2}{40.4} + \frac{(22-40.4)^2}{40.4} + \frac{(44-40.4)^2}{40.4} + \frac{(66-40.4)^2}{40.4} \\
= {} & \frac{268.96}{15.6} + \frac{31.36}{15.6} + \frac{12.96}{15.6} + \frac{.16}{15.6} + \frac{57.76}{15.6} \\
& + \frac{392.04}{35.8} + \frac{973.44}{35.8} + \frac{163.84}{35.8} + \frac{249.64}{35.8} + \frac{295.84}{35.8} \\
& + \frac{237.16}{59.6} + \frac{309.76}{59.6} + \frac{2246.76}{59.6} + \frac{243.36}{59.6} + \frac{876.16}{59.6} \\
& + \frac{2.56}{48.6} + \frac{54.76}{48.6} + \frac{158.76}{48.6} + \frac{750.76}{48.6} + \frac{31.36}{48.6} \\
& + \frac{108.16}{40.4} + \frac{.16}{40.4} + \frac{338.56}{40.4} + \frac{12.96}{40.4} + \frac{655.36}{40.4} = 195.36.
\end{aligned}$$

Substituting the value of X^2 and N in the Pearson's formula

$$C = \sqrt{\frac{X^2}{N+X^2}}$$

Where C represents the co-efficient of mean square contingency

We get,

$$C = \sqrt{\frac{195.36}{1000+195.36}} = 0.42.$$

Thus, the co-efficient of contingency is .4 which indicates that the association between the nationality of the subject and the nationality of the music preferred is possibly significant. From inspection of table, the contingency is positive which means that persons on the whole tended to prefer music of their own nationality. But there are exceptions, *e.g.*, English who on the whole appear to have preferred the music of German nationality.

PROBLEM 17.46. *Discuss the resemblance of stature of parent and offspring from the following :—*

Offspring	*Parent*				
	Very Tall	*Tall*	*Medium*	*Short*	*Total*
Very Tall	*20*	*30*	*20*	*2*	*72*
Tall	*14*	*125*	*85*	*12*	*236*
Medium	*3*	*140*	*165*	*125*	*433*
Short	*3*	*37*	*68*	*151*	*259*
Total	*40*	*332*	*338*	*290*	*1000*

SOLUTION. Let us take as our hypothesis the supposition that the two attributes, *viz.*, stature of parent and stature of offspring are independent. If this be true, the theoretical cell frequencies would be

Stature of Offspring	Stature of parent				Total
	Very Tall	Tall	Medium	Short	
Very Tall	2.9	23.9	24.3	20.9	72
Tall	9.4	78.4	79.8	68.4	236
Medium	17.3	143.8	146.4	125.5	433
Short	10.4	85.9	87.5	75.2	259
Total	40	332	338	290	1000

Substituting the observed and the theoretical frequencies in the formula

$$X^2 = S\left\{\frac{(f-f_1)^2}{f_1}\right\}$$

We get,

$$X^2 = \frac{(20-2.9)^2}{2.9} + \frac{(30-23.9)^2}{23.9} + \frac{(20-24.3)^2}{24.3} + \frac{(2-20.9)^2}{20.9}$$
$$+ \frac{(14-9.4)^2}{9.4} + \frac{(125-78.4)^2}{78.4} + \frac{(85-79.8)^2}{79.8} + \frac{(12-68.4)^2}{68.4}$$
$$+ \frac{(3-17.3)^2}{17.3} + \frac{(140-143.8)^2}{143.8} + \frac{(165-146.4)^2}{146.4} + \frac{(125-125.5)^2}{125.5}$$
$$+ \frac{(3-10.4)^2}{10.4} + \frac{(37-85.9)^2}{85.9} + \frac{(68-87.5)^2}{87.5} + \frac{(151-75.2)^2}{75.2}$$
$$= \frac{292.41}{2.9} + \frac{37.21}{23.9} + \frac{18.49}{24.3} + \frac{357.21}{20.9}$$
$$+ \frac{21.16}{9.4} + \frac{2171.59}{78.4} + \frac{27.04}{79.8} + \frac{3180.96}{68.4}$$
$$+ \frac{204.49}{17.3} + \frac{14.44}{143.8} + \frac{345.96}{146.4} + \frac{.25}{125.5}$$
$$+ \frac{54.76}{10.4} + \frac{2391.21}{85.9} + \frac{380.25}{87.5} + \frac{5745.64}{75.2} = 325.15$$

Substituting the value of X^2 and N in the Pearson's formula

$$C = \sqrt{\frac{X^2}{N+X^2}}$$

where C represents the co-efficient of Mean Square Contingency. We get,

$$C=\sqrt{\frac{325.15}{1000+325.15}}=.495$$

Thus, the co-efficient of contingency is .495 which indicates that the asso ation between the stature of parent and the stature of offspring is significa From inspection of the table, the contingency is positive which means that the is resemblance between the statures of the two.

REVISIONAL PROBLEMS–17

PROBLEM 17.47. Two research workers classified some people in incom groups on the basis of sampling studies. Their results are as follows :

Investigators	Poor	Middle class	Rich	Total
A	160	30	10	200
B	140	120	40	300
Total	300	150	50	500

Show that the sampling technique of at least one research worker defective.

(For 2 degrees of freedom, X^2 value at 1% level of significance = 9.21).

(*Ans* : $X^2=55.5$)

PROBLEM 17.48. The table given below shows the data obtained durin an outbreak of small pox :

	Attacked	Not attacked
Vaccinated	31	469
Not-Vaccinated	185	1315

Test the effectiveness of vaccination in preventing the attack from sma pox. Test your result with the help of X^2 at 5% level of significance. The valu of X^2 at 5% level of significance for one degree of freedom is 3.841.

(*Ans* : $X^2=14.65$, effective).

PROBLEM 17.49. The following information was obtained in a samp of 50 small general shops :

	Shops in		
	Towns	Villages	Total
Run by men	17	18	35
Run by women	3	12	15
Total	20	30	50

Can it be said that there are relatively more women owners of small general shops in villages than in town ?

(*Ans* : $X^2=3.57$.)

PROBLEM 17.50. Calculate the coefficient of contingency.

	A_1	A_2	A_3	A_4
B_1	11	6	2	1
B_2	5	12	15	8
B_3	—	2	3	15

(*Ans* : $C=0.59$).

PROBLEM 17.51. Two treatments were tried out in a control of a certain type of plant infection, and with the following results—

Treatment A : 200 plants examined and 24 found infected.

Treatment B : 200 plants examined and 9 found infected.

May we conclude that treatment B is superior to treatment A in controlling this type of infection ?

(*Ans* : $X^2=6.45$; $V=1$; $X^2\ 0.05=3.84$; yes).

PROBLEM 17.52. A normal distribution was fitted to the distribution of new business written by 100 insurance agents with the following results :

New business (000 Rs.)	10-20	20-30	30-40	40-50	50-60
Observed Frequency	10	20	33	22	15
Expected Frequency	5	22	32	25	12

Apply X^2 test.

(*Ans* : $X^2 = 1.43$.)

PROBLEM 17.53. The following figures show the distribution of digits in numbers chosen at random from a telephone directory :

Digit	0	1	2	3	4	5	6	7	8	9	Total
Frequency	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whether the digits may be taken to occur equally frequently in the directory.

(*Ans* : $x^2 = 58.542$; V = 9, x^2 0.05 = 16.919).

PROBLEM 17.54. A die is thrown 180 times with the following results :

No. turned up	1	2	3	4	5	6
Frequency	25	35	40	22	32	26

Test the hypothesis that the die is unbiased.

(*Ans* : $x^2 = 7.78$: V 5, x^2 0.05 = 11.07).

PROBLEM 17.55. A certain drug is claimed to be effective in curing cold. In an experiment with 160 persons, half of them were given the drug and half of them were given sugar pills. The patients' reaction to the treatment is recorded in the following table :

	Helped	Harmed	No effect
Drug	52	10	18
Sugar Pill	44	10	26

Test the hypothesis that the drug is no better than sugar pills for curing colds at 5% level.

(*Ans* : $x^2 = 2.12$).

PROBLEM 17.56. An experiment was conducted to test the efficacy of chloromycin in checking typhoid. In a certain hospital chloromycin was given to 300 out of 400 patients suffering from typhoid. The number of typhoid cases is as follows :

	Typhoid	No Typhoid	Total
Chloromycin	40	260	300
No Chloromycin	60	40	100
	100	300	400

With the help of x^2 test find the effectiveness of chloromycin.

(*Ans.* $x^2 = 87.1$).

Sampling of Variables—Small Samples

18

In this chapter various tests of significance have been applied to small samples. First of all problems have been done to test the significance of the Mean in small samples. In most of these problems "*t*" test has been applied. The following formulae which have been explained at places where they are used have been made use of :—

(*i*) *To test the significance of the mean*

$$t=\frac{\overline{x}-m}{\sigma}\sqrt{n}$$

(*ii*) *To test the significance of the difference of two sample means*

$$t=\frac{x_1-x_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\text{ where}$$

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+(x_2-\bar{x}_2)^2}{n_1+n_2^{2}}}$$

(*iii*) *To test the significance of co-efficient of correlation*

$$t=\frac{\sqrt{n-2}}{\sqrt{1-r^2}}\times n$$

Correlation can also be tested by '*z*' test

$$z=\log_{10}\frac{1+r}{1-r}\times 1.1513$$

Standard error of $z=\frac{1}{\sqrt{n-3}}$

(*iv*) *To test the significance of the difference between two sample co-efficients of correlation.*

Co-efficients of correlation are first converted in Z by the formula used above and then the standard error of the difference of two *z's* is calculated by the following formula

$$s.e.\ z_1\text{-}z_2=\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}$$

At this stage "*t*" test can also be applied.

$$t=\frac{Z_1-Z_2}{s.e.Z_1\text{-}Z_2}$$

In all the above cases the calculated value of "t" is compared with the table value and the conclusions are drawn.

SIGNIFICANCE OF MEAN

PROBLEM 18.1. *Ten individuals are chosen at random from a population and their heights are found to be, in cms. 163,163,164,165,166,169,169,170,170,171. In the light of these data, discuss the suggestion that the mean height in the universe is 165 cms.*

SOLUTION. Calculation of the average height and its standard deviation.

No. of individuals	*Height in centimetres (cm)*	*Deviation from the average (167) (d)*	*Square of deviation (d^2)*
1	163	—4	16
2	163	—4	16
3	164	—3	9
4	165	—2	4
5	166	—1	1
6	169	+2	4
7	169	+2	4
8	170	+3	9
9	170	+3	9
10	171	+4	16
$n=10$	$\Sigma m=1670$		$\Sigma d^2=88$

Average height of the sample $=\frac{\Sigma m}{n}=\frac{1670}{10}=167$ cms.

Standard deviation of the height of the sample is

$$\sigma=\sqrt{\frac{\Sigma d^2}{n-1}}=\sqrt{\frac{88}{10-1}}=3.13 \text{ cms.}$$

Substituting these values in the formula.

$$t=\frac{\bar{x}-m}{\sigma}\sqrt{n}$$

Where x represents the mean of the sample, m the mean of the aniverse, σ the standard deviation of the sample and n the number of observations.

We get,

$$t=\frac{167-165}{3.13}\sqrt{10}=2.02$$

The number of degrees of freedom is $n-1=9$.

The value of t for 9 degrees of freedom at 5% level of significance is 2.262. The calculated value of t is 2.02 and is less than the table value so at 5% level of significance this error could have arisen due to fluctuations of sampling and it can be said that the mean height in the universe is 165 cms.

PROBLEM 18.2. *Find "Students" "t" for the following variate values in a sample of 10 :—6,—4,—3,—2,—2,0,1,1,3,5 ; taking m to be zero.*

SOLUTION. Calculation of the mean and the standard deviation the sample.

Values of the items (*m*)	*Deviation from the average* (—.7) (*d*)	*Square of deviation* (d^2)
—6	—5.3	28.09
—4	—3.3	10.89
—3	—2.3	5.29
—2	—1.3	1.69
—2	—1.3	1.69
0	.7	.49
1	1.7	2.89
1	1.7	2.89
3	3.7	13.69
5	5.7	32.49
$\Sigma m = -7$		$\Sigma d^2 = 100.10$

Mean of the sample

$$\bar{x} = \frac{\Sigma m}{n} = \frac{-7}{10} = -.7$$

Standard deviation of the sample

$$\sigma = \sqrt{\frac{\Sigma d^2}{n-1}} = \sqrt{\frac{100.1}{10-1}} = 3.34$$

Substituting the calculated and the given values in the formula on the n hypothesis

$$t = \frac{\bar{x}-m}{\sigma}\sqrt{n}$$

We get,

$$t = \frac{-.7-0}{3.34}\sqrt{10}$$

$$= -0.662.$$

PROBLEM 18.3. *The heights of ten men taken at random from a norma population were 163,164,165,167,168,169,169,170,172, cms. respectively. Do thes data support the hypo.hesis of a population mean of 166.1 cms. ?*

SOLUTION. Sample Mean $=\bar{x}=$ 1507 cms. 167.4 cms.

Sample variance $\sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

$$= \frac{21.16+19.36+11.56+5.76+0.16+0.36+2.56+2.56+6.76}{9}$$

$$= \frac{70.24}{9} = 7.80$$

$$\therefore \quad \sigma = \sqrt{7.80} = 2.793$$

$$\text{Now, } t=\frac{(\bar{x}-m)\sqrt{n}}{\sigma}=\frac{167.4-166.1\sqrt{10}}{2.793}=\frac{1.3\times 3.162}{2.793}$$

$$=\frac{4.1106}{2.796}=1.47;\ \text{Degrees of freedom}=10-1=9$$

From the table of (t) we find that for 9 degrees of freedom, $t=1.83$ is significant at 0.10 level. The observed value of t is less than this and, therefore, not at all significant even at 10% level.

The hypothesis is consistent with the data.

PROBLEM 18.4. *A coin is tossed 16 times and the number of heads noted. This experiment is made 10 times. The number of heads shown are 3,5,4,8,10,9,2,7,4,8 respectively. Is it likely that the coin is biased ?*

SOLUTION. If our hypothesis is that the coin is unbiased, in that case the mean $m=8$.

$$\text{Sample Mean } x=\frac{3+5+4+8+10+9+2+7+4+8}{10}$$

$$=\frac{60}{10}=6$$

$$\text{Sample Variance } \sigma^2=\frac{9+1+4+4+16+9+16+1+4+4}{10-1}$$

$$=\frac{68}{9}=7.556$$

$$\therefore\ \sigma=2.749$$

$$t=\frac{(\bar{x}-m)\sqrt{n}}{\sigma}=\frac{(6-8)\sqrt{10}}{2.749}=\frac{2\times 3.162}{2.749}=\frac{6.324}{2.749}=2.30$$

Degrees of freedom$=10-1=9$

From table of t we find that for 9 degrees of freedom $t=2.26$ is significant for 05 level. The observed value of t is greater than this and, therefore, significant. The hypothesis is doubtful, *i.e.*, the coin is biased.

PROBLEM 18.5. *Ten individuals are chosen at random from a population and their heights are found to be in cms. 163,164,166,167,168,169,170,170,170, and 171. In the light of these data discuss the suggestion that the mean height in population is 166 cms.*

SOLUTION. In the above question the value of the mean of the sample or $\bar{x}$ would come to 167.8 cms. and the value of standard deviation or σ would come to 2.75 cms.

As such the value of

$$t=\frac{\bar{x}-m}{\sigma}\sqrt{n}\ =\frac{67.8-66}{2.75}\sqrt{10}=2.08$$

At 5% level of significance the value or t for 9 degrees of freedom is 2.262. The calculated value is less than this and as such the difference may be due to sampling fluctuations and it can be said that the mean height of the universe is 166 cms.

PROBLEM 18.6. *A machine which produces mica insulating washers for use in electrical devices is set to turn out washers having a thickness of 10 millimetres. (1 mm $=0.0001$ metre). A sample of 10 washers has an average thickness of 9.52 mm. with a standard deviation of 0.60 mm. Find out the significance of such a deviation.*

SOLUTION. Substituting the given values in the formula

$$t=\frac{\bar{x}-m}{\sigma}\sqrt{n}$$

Where x represents the mean of the sample ; m the mean of the universe ; σ the standard deviation of the sample ; n the number of observations in the sample.

We get,

$$t=\frac{9.52-10}{0.60}\sqrt{10}=2.5296 \text{ or } -2.55 \text{ Approx.}$$

The number of degrees of freedom is 10—1, or 9. The value of t for 9 degrees of freedom at 5% level of significance is 2.262. The calculated value is greater than this. Hence, the deviation is significant.

PROBLEM 18.7. *The mean of a random sample of 26 items from a normal population is 42.5 units. The 'sum of squares' from this mean is 250. Is a population mean of 45 units likely ? What are the (i) 99%, (ii) 98%, and (iii) 95% fiducial limits af the population mean ?*

SOLUTION. Sample Mean : $\bar{x}=42.5$

Given : $n=26$ and $\Sigma\,(\bar{x}-x)^2=250$

$$\therefore \text{ Sample Variance } \sigma^2=\frac{250}{26-1}=\frac{250}{25}=10$$

$$\therefore \sigma=\sqrt{10}=3.162$$

$$\therefore t=\frac{(\bar{x}-m)\sqrt{n}}{\sigma}=\frac{(42.5-45)\sqrt{26}}{\sqrt{10}}=-2.5\sqrt{2.6}$$

$$=-2.5\times1.612=-4.0300=-4.03$$

Degrees of freedom$=26-\,-1=25$

From table of t we find that for 25 degrees of freedom $t=3.73$ is significant at .001 level of significance. The observed value of t is greater than this and, therefore, very highly significant. It is very unlikely that it is a sample from a population the mean of which is 45 units.

(*ii*) Fiducial limits of the population mean :

For 25 degrees of freedom we find :
99% or .01 level value of $t=2.79$
98% or .02 level value of $t=2.49$
95% or .05 level value of $t=2.06$

$$\therefore \text{ 99\% fiducial limits of population mean }=42.5\pm\frac{3.162}{\sqrt{26}}\times2.79$$

$$=42.5\pm1.73 \text{ or } 44.23 \text{ and } 40.77$$

98% fiducial limits of population mean $=42.5 \pm \frac{3.162}{\sqrt{26}} \times 2.49$

$=42.5 \pm 1.54$ or 44.04 and 40.96

95% ,, ,, ,, $=42.5 \pm \frac{3.162}{\sqrt{26}} \times 2.06$

$=42.5 \pm 1.28$ or 43.78 and 41.22

PROBLEM 18.8. *A certain stimulus administered to each of 12 patients ulted in the following increase of blood pressure ; 5,2,8,–1,3,0,6,–2,1,5,0,4. Can t be concluded that the stimulus will be in general accompanied by an increase in lood pressure ?*

SOLUTION. Calculation of the mean and the standard deviation of the ample.

Increase in blood pressure of 12 patients (*m*)	*Deviation from the as. av.* (*3*) (*dx*)	*Square of deviation* (d^2x)
5	2	4
2	–1	1
8	5	25
–1	–4	16
3	0	0
0	–3	9
6	3	9
–2	–5	25
1	–2	4
5	2	4
0	–3	9
4	1	1
$n=12$	$\Sigma dx=-5$	$\Sigma d^2x=107$

Mean of the sample

$$\bar{x}=x+\frac{\Sigma dx}{n}=3+\frac{-5}{12}=2.58.$$

Standard deviation of the sample

$$\sigma=\sqrt{\frac{\Sigma d^2x}{n}-\left(\frac{\Sigma dx}{n}\right)^2}=\sqrt{\frac{107}{12}-\left(\frac{-5}{12}\right)^2}=2.96$$

We take for our hypothesis the supposition that the stimulus does not result in an increase in blood pressure, that is $m=0$.

Now, substituting these values in the formula

$$t=\frac{\bar{x}-m}{\sigma}\sqrt{n-1}$$

We get,

$$t=\frac{2.58-0}{2.96}\sqrt{-11}$$
$$=2.89$$

The number of degrees of freedom is 12—1, or 11. The value of t for 11 degrees of freedom at 5% level of significance is 2.201 which is less than the calculated value of t. So, the deviation is significant which means that the hypothesis is incorrect. Hence it can be concluded that the stimulus will be in general accompanied by an increase in blood pressure.

Note :—The original formula for the calculation of

$t=\frac{\bar{x}-m}{\sigma}\sqrt{n}$. In small samples while calculating standard deviation the formula used is $\sigma=\sqrt{\frac{\Sigma fd^2}{n-1}}$ but in this case we have used the formula $\sqrt{\frac{\Sigma fd^2}{n}}$ in order to adjust the difference we have modified the formula for the calculation of t by using $\sqrt{n-1}$ instead of $\sqrt{\,}\sqrt{n}$.

PROBLEM 18.9. *What is students "t" distribution ? Explain its use in small samples.*

Ten individuals are chosen at random from a population and their heights are found to be, in cms., 163, 163, 166, 167, 168, 169, 170, 170, 171, 171. In the light of these data, discuss the suggestion that the mean height in the universe is 166 cms.

Given that V=9. P=.947 for t=1.8

P=.955 for t=1.9

SOLUTION. Calculation of the average height and its standard deviation.

No. of individuals	*Height in cms. (cm)*	*Deviation from the average 167.8 (d)*	*Square of deviation* $(d)^2$
1	163	—4.8	23.04
2	163	—4.8	23.04
3	166	—1.8	3.24
4	167	— .8	.64
5	168	+ .2	04
6	169	+1.2	1.44
7	170	+2.2	4.84
8	170	+2.2	4.84
9	171	+3.2	10.24
10	171	+3.2	10.24
$n=10$	$\Sigma m=1678$		$\Sigma d^2=81.60$

Average height of the sample $= \frac{\Sigma m}{n} = \frac{1678}{10} = 167.8$cms.

Standard deviation of the height of the sample is

$$\sigma = \sqrt{\frac{\Sigma d^2}{n-1}} = \sqrt{\frac{81.6}{10-1}} = 3.01 \text{ cms.}$$

Substituting these values in the formula

$$t = \frac{\bar{x}-m}{\sigma}\sqrt{n}$$

Where $\bar{x}$ represents the mean of the sample, m the mean of the universe, σ the standard deviation of the sample and n the number of observations.

We get,

$$t = \frac{167.8-166}{3.01}\sqrt{10} = 1.89$$

The number of degrees of freedom is $n-1=9$.

The value of t for 9 degrees of freedom at 5% level of significance is 2.262. The calculated value of t is 1.89 and is less than the table value. So at 5% level of significance this error could have arisen due to fluctuations of sampling and it can be said that the mean height in the universe is 166 cms.

PROBLEM 18.10. *The sleep of 10 patients was measured for the effects of two soporific drugs referred to in the following table as (1) and (2). From the data given below show that there is a significant difference between the effects of the two drugs.*

Additional hours of sleep gained by use of soporifics (1) and (2)

Patients	*(1)*	*(2)*	*Difference (2) (1)*
1	*+ .7*	*1.9*	*1.2*
2	*−1.6*	*.8*	*2.4*
3	*− .2*	*1.1*	*1.3*
4	*−1.2*	*.1*	*1.3*
5	*− .1*	*− .1*	*.0*
6	*3.4*	*4.4*	*1.0*
7	*3.7*	*5.5*	*1.8*
8	*.8*	*1.6*	*.8*
9	*.0*	*4.6*	*4.6*
10	*2.0*	*3.4*	*1.4*
Mean	*.75*	*2.33*	*1.58*
σ	*1.70*	*1.90*	*1.17*

SOLUTION. We proceed on the null hypothesis, *i.e.*, on the supposition that there is no difference in the effects of the two drugs, that is, the mean of the differences or $m=0$.

Now, substituting the given values in the formula,

$$t = \frac{(\bar{x}-m)}{\sigma}\sqrt{n-1}$$

We get,

$$t=\frac{1.58-0}{1.17}\sqrt{10-1}=4.05.$$

The number of degrees of freedom 10--1, or 9. The value of t for 9 degre of freedom at 5% level of significance is 2.262 which is considerably smaller tha the calculated value. Hence the hypothesis is incorrect and there is a significa difference between the effects of the two drugs which is in favour of the second.

Note :—As in an earlier question, here also σ has been calculated by usi the formula $\sqrt{\frac{d^2}{n}}$ instead of $\sqrt{\frac{fd^2}{n-1}}$ and hence while calculating t adjustme has been made by writing $\sqrt{n-1}$ in place of $\sqrt{n}$.

PROBLEM 18.11. *A sleeping draught and a neutral control preparation we tested in turn (in random order) on 10 patients at a hospital and gave the followi results :*

No. of hours of sleep			*No. of hours of sleep*		
Patients	***Drug***	***Control***	***Patients***	***Drug***	***Contro***
1	*10.6*	*8.6*	*6*	*10.2*	9.0
2	*7.5*	*7.3*	*7*	*7.1*	6.5
3	*9.0*	*9.4*	*8*	*9.7*	7.9
4	*.5.4*	*5.1*	*9*	*8.5*	8.7
5	*6.1*	*5.4*	*10*	*7.9*	6.9

Calculate the appropriate value of "t" to test whether the drug has any grea effect than the neutral control preparation.

SOLUTION. We have to test whether the drug has any greater effect. there is no such effect, the expected mean difference in the result is zero. Thi our hypothesis.

Calculation of t

Patients	*No. of hours of sleep*		***Difference (d)***	$(d)^2$
	Drug	***Control***		
1	10.6	8.6	2.0	4.00
2	7.5	7.3	0.2	0.04
3	9.0	9.4	—0.4	0.16
4	5.4	5.1	0.3	0.09
5	6.1	5.4	0.7	0.49
6	10.2	9.0	1.2	1.44
7	7.1	6.5	0.6	0.36
8	9.7	7.9	1.8	3.24
9	8.5	8.7	—0.2	0.04
10	7.9	6.9	1.00	1.00
			7.8	10.86
			—0.6	
			7.2	

$$\bar{d} = \frac{7\cdot2}{10} = 0.72 \; ; \; \Sigma(\bar{d}-d)^2 = \Sigma d^2 - \frac{1}{n}(\Sigma d)^2$$

$$= 10.86 - \frac{51.84}{10} = 10.86 - 5.184 = 5.676$$

$$\therefore \sigma^2 = \frac{\Sigma(\bar{d}-d)^2}{n-1} = \frac{5.676}{9} \therefore \sigma = \frac{2.382}{3} = 0.794$$

$$\therefore t = \frac{(\bar{d}-0)\sqrt{n}}{\sigma} = \frac{0.72 \times 3.162}{0.794} = 2.867$$

Degrees of freedom $= 10 - 1 = 9$

For 9 degrees of freedom, $t = 2.82$ is significant at .02 level, the observed value of t is greater than this value and, therefore, the difference is significant.

The results show that the drug has greater effect.

PROBLEM 18.12. *Eleven school boys were given a test in drawing. They were given a month's further tuition and a second test of equal difficulty was held at the end of it. Do the marks give evidence that the students have benefited by extra coaching ?*

Boys	*Marks 1st Test*	*Marks 2nd Test*
1	23	24
2	20	19
3	19	22
4	21	18
5	18	20
6	20	22
7	18	20
8	17	20
9	23	23
10	16	20
11	19	17

SOLUTION. Calculation of the mean and the standard deviation of the differences between the marks of the two tests :

Boys	*Marks 1st Test*	*Marks 2nd Test*	*Differences of marks (2nd Test — 1st Test)* (*m*)	*Deviation from the average (1)* (*d*)	*Square of deviation* (d^2)
1	23	24	+1	0	0
2	20	19	−1	−2	4
3	19	22	+3	2	4
4	21	18	−3	−4	16
5	18	20	+2	1	1
6	20	22	+2	1	1
7	18	20	+2	1	1
8	17	20	+3	2	4
9	23	23	0	−1	1
10	16	20	+4	3	9
11	19	17	−2	−3	9
			$\Sigma m = 11$		$\Sigma d^2 = 50$

Mean of the differences

$$\bar{x} = \frac{\Sigma m}{n} = \frac{11}{11} = 1.$$

Standard deviation of the differences

$$\sigma = \sqrt{\frac{\Sigma d^2}{n-1}} = \sqrt{\frac{50}{11-1}} - 2.24$$

Now, let us proceed on the null hypothesis, *i.e.*, on the supposition that the students have not been benefited by the extra coaching, that is, the mean of the differences between the marks of the two tests is zero, or $m=0$.

Substituting these values in the formula,

$$t = \frac{\bar{x} - m}{\sigma}\sqrt{n}$$

We get,

$$t = \frac{1-0}{2.24}\sqrt{11} = 1.482.$$

The number of degrees of freedom is 11—1, or 10.

The value of t for 10 degrees of freedom at 5% level of significance is 2.228. The calculated value is considerably less than this value. Hence, the hypothesis is correct and the marks do not give any evidence that the students have benefited by the extra coaching.

PROBLEM 18.13. *A group of 8 psychology students were tested for their ability to remember certain material and their scores (number of items remembered) were as follows :*

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
(*1*)	*19*	*14*	*13*	*16*	*19*	*18*	*16*	*17*

They were then given special training purporting to improve memory and were re-tested after a month. Their scores were then :

(2)	*26*	*20*	*17*	*21*	*23*	*24*	*21*	*18*

A control group of seven students were also tested and re-tested after a month, but was not given any special training. The scores in the two tests were :

	J	*K*	*L*	*M*	*N*	*O*	*P*
(*1*)	*21*	*19*	*16*	*22*	*18*	*20*	*19*
(*2*)	*21*	*23*	*16*	*24*	*17*	*17*	*16*

Compare the change in each of the two groups by calculating and test whether there is significant evidence to show the value of the special training.

SOLUTION. If special training has no effect, the expected mean difference is zero. This is our hypothesis.

First group :	A	B	C	D	E	F	G	H	Total
1st test	19	14	13	16	19	18	16	17	
2nd Test (after special training)	26	20	17	21	23	24	21	18	

Differences (d) :	7	6	4	5	4	6	5	1 = 38
(d^2)	49	36	16	25	16	36	25	1 = 204

$$\overline{d}=\frac{38}{8}=4.75$$

$$\Sigma(\overline{d}-d)^2=\Sigma d^2-\frac{1}{n}(\Sigma d)^2=204-\frac{1444}{8}=204-180.5=23.5$$

$$\sigma^2=\frac{\Sigma(\overline{d}-d)^2}{n-1}=\frac{23.5}{7}=3.3571$$

$$\therefore\ \sigma=1.83$$

Now, $t=\dfrac{(\overline{d}-0)\sqrt{n}}{\sigma}=\dfrac{4.75\ \sqrt{8}}{1.83}=\dfrac{4.75\times 2.828}{1.83}=\dfrac{13.43}{1.83}=7.33$

Degrees of freedom$=8-1=7$

For 7 degrees of freedom, $t=5.41$ is significant at .001 level. The observed value of t is greater than this and, therefore, very highly significant.

Second group : J	K	L	M	N	O	P	Total
1st Test : 21	19	16	22	18	20	19	
2nd Test : 21	23	16	24	17	17	16	
Difference (d) : 0	4	0	2	-1	-3	-3	$=-1$
d^2 : 0	16	0	4	1	9	9	$=39$

$$\Sigma d^2-\frac{1}{n}(\Sigma d)^2=39-\frac{1}{7}=38\frac{6}{7};\ \sigma^2=\frac{272}{7\times 6}=\frac{272}{42};\ \therefore\ \sigma=2.54$$

$$\therefore t=\frac{0.1667\times\sqrt{7}}{2.54}=\frac{0.1667\times 2.646}{2.54}=0.173$$

Degrees of freedom$=7-1=6$

For 6 degrees of freedom, $t=0.72$ is significant at 0.50 level. The observed value of t is less than this and, therefore, not at all significant.

A comparison of these two values of t shows that special training had some value.

PROBLEM 18.14. *The following table shows the average yearly mortality in utomobile accidents per 100,000 population in 5 cities of a country for two three-year eriods.*

Cities	*Rate per 100,000 for 1959-61 period*	*Rate per 100,000 for 1970-72 period*
A	*26.6*	*28.5*
B	*18.1*	*21.3*
C	*15.0*	*17.7*
D	*27.6*	*35.8*
E	*20.9*	*25.4*

Determine whether or not there was a significant increase in the mortality in utomobile accidents from the first period to the second.

SOLUTION. Calculation of the mean and the standard deviation of the differences between the mortality rates of the two periods.

Cities	*Rate per 100,000 for 1959-61 period*	*Rate per 100,000 for 1970-72 period*	*Differences* (m)	*Deviation from the mean (4.1)* (d)	*Square of deviation* $(d)^2$
A	26.6	28.5	1.9	−2.2	4.84
B	18.1	21.3	3.2	− .9	.81
C	15.0	17.7	2.7	−1.4	1.96
D	27.6	35.8	8.2	+4.1	16.81
E	20.9	25.4	4.5	+ .4	.16
			$\Sigma m=20.5$		$\Sigma d^2=24.58$

Mean of the differences

$$\bar{x}=\frac{\Sigma m}{n}=\frac{20.5}{5}=4.1$$

Standard deviation of the differences

$$\sigma=\sqrt{\frac{\Sigma d^2}{n-1}}=\sqrt{\frac{24.58}{5-1}}=2.47$$

On the null hypothesis we presume that no increase has taken place in the mortality in automobile accidents from the first period to the second, that is, the true mean of the difference or $m=0$.

Substituting these values in the formula.

$$t=\frac{\bar{x}-m}{\sigma}\sqrt{n}$$

We get,

$$t=\frac{4.1-0}{2.27}\sqrt{5}=3.7.$$

The number of degrees of freedom is 5—1, or 4.

The value of t for 4 degrees of freedom at 5% level of significance is 2.776 which is less than the calculated value, so the deviation from the hypothesis is significant. Hence the hypothesis is incorrect and there seems to be a significant increase in the mortality in automobile accidents from the first period to the second.

PROBLEM 18.15. *A farmer grows crops on two fields, A and B. On A he puts Rs. 100 worth of manure per acre and on B Rs. 200 worth. The net returns per acre, exclusive of the cost of manure, on the two fields in five years are :*

Year	*Field A, Rs. per Acre*	*Field B, Rs. per Acre*
1	*234*	*236*
2	*228*	*233*
3	*242*	*248*
4	*237*	*238*
5	*244*	*250*

Other things being equal, discuss the question whether it is likely to pay the farmer to continue the more expensive dressing.

SOLUTION. Calculation of the mean and the standard deviation of the differences in the net yields of the two farms·

Year	Field A Net yield in rupees per acre	Field B Net yield in rupees per acre	Differences (B)—(A) (m)	Deviation from the average (4) (d)	Square of deviation (d^2)
1	234	236	2	—2	4
2	228	233	5	+1	1
3	242	248	6	+2	4
4	237	238	1	—3	9
5	244	250	6	+2	4
n=5			$\Sigma m=20$		$\Sigma d^2=22$

Mean of differences of the net yields is

$$\overline{x}=\frac{\Sigma m}{n}=\frac{20}{5}=4 \text{ rupees.}$$

Standard deviation of the differences of the net yields.

$$\sigma=\sqrt{\frac{\Sigma d^2}{n-1}}=\sqrt{\frac{22}{5-1}}=2.345 \text{ rupees.}$$

Now, we take as our hypothesis, the supposition that the dressing has got no effect on net yields, that is, the true mean of the difference, or m—0.

Substituting these values in the formula

$$t=\frac{\overline{x}-m}{\sigma}\sqrt{n} \quad \text{we get,}$$

$$t=\frac{4-0}{2.345}\sqrt{5}=3.814.$$

The number of degrees of freedom is 5 − 1, or 4.

The value of t for 4 degrees of freedom at 5% level of significance is 2.776 which is smaller than the calculated value, 3.814. So, the hypothesis is wrong and the increased returns on the better manured plot seem to be real. Hence, it can be concluded that other things remaining the same, it is likely to pay the farmer to continue the more expensive dressing.

PROBLEM 18.16. *Why should there be different formulae for testing significance of difference in means when the samples are (a) 'small' and (b) 'large' ?*

The yields of two types, "Type 17" and "Type 51" of grams in Kilograms per plot of the same size at 6 replications are given below. What comments would you make on the differences in the mean yields ? You may assume that if there be 5 degrees of freedom and P=0.2, t is approximately 1.476.

Replication	Yield in Kilograms 'Type 17'	Yield in Kilograms 'Type 51'
1	20.50	24.86
2	24.60	26.39
3	23.06	28.19
4	29.98	30.75
5	30.37	29.97
6	23.83	22.04

SOLUTION. Calculation of the mean and the standard deviation of the differences in the yields of two types of gram.

Replication	*Yield in Kilograms 'Type 17'*	*Yield in Kilograms 'Type 51'*	*Differences* (m)	*Deviation from the average of the difference (1.643)* (d)	*Square of deviation* (d^2)
1	20.50	24.86	4.36	+2.717	7.383
2	24.60	26.39	1.79	+0.147	0.022
3	23.06	28.19	5.13	+3.487	12.15
4	29.98	30.75	0.77	−0.873	0.762
5	30.37	29.97	−0.40	−2.043	4.174
6	23.83	22.04	−1.79	−3.433	11.79
N=6			$\Sigma m=9.86$		$\Sigma d^2=$ 36.281

Mean of the differences of the yield is

$$\bar{x}=\frac{\Sigma m}{n}=\frac{9.86}{6}=1.643 \text{ Kilograms}$$

Standard deviation of the difference of the yields

$$\sigma=\sqrt{\frac{\Sigma d^2}{n-1}}=\sqrt{\frac{36.281}{6-1}}=2.7 \text{ Kilograms}$$

Now, we proceed with the null hypothesis, that the difference in the types has got no effect on yields, that is, the true mean of the difference, or m=0.

Substituting these values in the formula

$$=\frac{\bar{x}-m}{\sigma}\sqrt{n}$$

We get,

$$t=\frac{1.643-0}{2.7}\sqrt{6}\quad=1.486$$

The number of degrees of freedom is 6—1, or 5.

The value of t for 5 degrees of freedom at 2% level of significance is 1.476. The calculated value of it is greater than this value. Hence it sets doubt on our hypothesis and it seems that at this level of significance there is a difference in yields as corresponding to difference in the types of gram.

PROBLEM 18.17. *Two laboratories carry out independent estimates of fat content for ice-cream made by a certain firm. A sample is taken from each batch, halved, and the separate halves sent to the two laboratories. They obtain the following results.*

Percentage fat content in ice-cream.

Batch No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Lab. A	*7*	*8*	*7*	*3*	*8*	*6*	*9*	*4*	*7*	*8*
Lab. B	*9*	*8*	*8*	*4*	*7*	*7*	*9*	*6*	*6*	*6*

Is the testing reliable ?

SOLUTION. If the testing is reliable the mean of the difference in results should not significantly differ from zero.

We shall calculate the mean and the standard deviation of the difference of results.

Calculation of the mean and standard deviation of results of the tests.

Difference of results (*Lab. B—Lab. A*)	*Deviation from mean* (.3) (d)	(d^2)
+2	+1.7	2.89
0	− .3	.09
+1	+ .7	.49
+1	+ .7	.49
−1	−1.3	1.69
+1	+ .7	.49
0	− .3	.09
+2	+1.7	2.89
−1	−1.3	1.69
−2	−2.3	5.29
+3		$\Sigma d^2$16.10

Mean of the difference is

$$\bar{x} = \frac{\Sigma m}{n} = \frac{3}{10} = .3$$

Standard deviation of the differences

$$\sigma = \sqrt{\frac{\Sigma d^2}{n-1}} = \sqrt{\frac{16.1}{10-1}} = 1.33$$

Taking the true mean as 0 and substituting the figures in the formula

We get,

$$t = \frac{\bar{x}-m}{\sigma}\sqrt{n}$$

We get,

$$t = \frac{.3-0}{1.33}\sqrt{10} = .71.$$

The value of t for 9 degrees of freedom at 5% level is 2.57. The calculated value is much less than this and as such the difference is insignificant and the tests have been reliable.

PROBLEM 18.18. *Find out the reliability of the sample mean of the following data :—*

Breaking strength of 10 specimens of 1.04 cms. diameter harddrawn copper wire.

Specimen	*Breaking-strength in kilograms*	*Specimen*	*Breaking-strength in Kilograms*
1	*578*	*6*	*570*
2	*572*	*7*	*570*
3	*570*	*8*	*572*
4	*568*	*9*	*526*
5	*572*	*10*	*584*

SOLUTION. Calculation of the sample mean and its standard deviation.

Specimen	*Breaking strength in kilograms* (m)	*Deviation from as. av.* (572) (dx)	*Square of Deviation* (dx^2)
1	578	$+6$	30
2	572	0	0
3	570	-2	4
4	568	-4	61
5	572	0	0
6	570	-2	4
7	570	-2	4
8	572	0	0
9	596	$+24$	576
10	584	$+12$	144
$n=10$		$\Sigma dx=32$	$\Sigma d^2x=784$

Mean breaking strength

$$\overline{x}=x+\frac{\Sigma dx}{n}=572+\frac{32}{10}=575.2$$

Standard deviation of the breaking strength

$$\sigma=\sqrt{\frac{\Sigma d^2x}{n}-\left(\frac{\Sigma dx}{n}\right)^2}=\sqrt{\frac{784}{10}-\left(\frac{32}{10}\right)^2}$$

$= 8.26$ kilograms.

Standard error of the sample mean

$$\sigma \text{ mean}=\frac{\sigma}{\sqrt{n-1}}=\frac{8.26}{\sqrt{10-1}}=2.75 \text{ kilograms.}$$

Thus the sample mean of the breaking strength is 575.2 kilograms and its standard error is 2.75 kilograms.

PROBLEM 18.19. *The following data relate to the stature and shoulder height of 10 boys taken once in the morning at 6 a. m. and again at night at about 10 p. m. on the same day. Examine by appropriate tests whether the morning averages differ significantly from the night averages. It is believed that certain portions increase in length in a relaxed position of the body so that the measurements in the morning immediately after we get up from bed would have higher value. Can you indicate which portion of the body, if any, does so on the basis of measurements provided ?*

Serial No. of days	*Stature*		*Shoulder Height*	
	Morning	*Night*	*Morning*	*Night*
1	*177.3*	*176.1*	*164.6*	*164.9*
2	*165.9*	*165.1*	*154.2*	*154.3*
3	*174.1*	*173.1*	*161.1*	*161.6*
4	*170.4*	*169.3*	*158.8*	*158.2*
5	*166.6*	*165.9*	*153.0*	*152.9*
6	*167.9*	*167.4*	*156.8*	*156.8*
7	*163.3*	*162.6*	*152.3*	*151.9*
8	*160.4*	*159.1*	*146.2*	*146.8*
9	*158.4*	*157.9*	*147.4*	*146.9*
10	*160.7*	*159.9*	*148.5*	*148.0*

SOLUTION. The data occur in pairs and, therefore, we apply 'The Difference Test'. Our hypothesis is that there is no significant difference between the heights. In that case in the expected mean differece is zero.

Serial No. of boys	Stature		Difference	
	Morning	Night	(d_1)	(d_1^2)
1	177.3	176.1	1.2	1.44
2	165.9	165.1	0.8	0.64
3	174.1	173.1	1.0	1.00
4	170.4	169.3	1.1	1.21
5	166.6	165.9	0.7	0.49
6	167.9	167.4	0.5	0.25
7	163.3	162.6	0.7	0.49
8	160.4	159.1	1.3	1.69
9	158.4	157.9	0.5	0.25
10	160.7	159.9	0.8	0.64
			8.6	8.10

$$\bar{d}_1=\frac{8.6}{10}=0.86$$

$$\Sigma(\bar{d}_1-d_1)^2=\Sigma d_1^2-\frac{(\Sigma d_1)^2}{n}=8.10-\frac{73.96}{10}=8.10-7.396=0.704$$

$$\sigma_1^2=\frac{0.704}{9}=0.0782 \quad \therefore \quad \sigma_1=\sqrt{0.782}=0.28$$

$$t=\frac{(\bar{d}_1-0)\sqrt{n}}{\sigma_1}=\frac{0.86\sqrt{10}}{0.28}=9.71$$

For 9 degrees of freedom, $t=4.78$ is significant at .001 level of significance. Observed value of t is greater than this value and, therefore, highly significant. The data suggest that measurements in the morning have higher values—so far as stature is concerned.

Morning	Night	Shoulder Height Difference		
			(d_2)	(d_2^2)
164.6	164.9		-0.3	0.09
154.2	154.3		-0.1	0.01
161.1	161.6		-0.5	0.25
158.8	158.2		0.6	0.36
153.0	152.9		0.1	0.01
156.8	156.8		0.0	0.00
152.3	151.9		0.4	0.16
146.2	146.8		-0.6	0.36
147.4	146.9		0.5	0.25
148.5	148.0		0.5	0.25
			0.6	+1.74

$$\overline{d_2}=\frac{0.6}{10}=0.06$$

$$\Sigma(\overline{d_2}-d_2)^2=1.74-\frac{0.36}{10}=1.74-0.036=1.704$$

$$\sigma_2^2=\frac{1.704}{9}=0.1893 \therefore \sigma_2=\sqrt{0.1893}=0.435$$

$$t=\frac{.06\times 3.162}{0.435}=\frac{.18972}{0.435}=0.436$$

For 9 degrees of freedom, $t=0.70$ is significant at .50 level. Observed value of t is less than this and, therefore, not at all significant. The data support the hypothesis that so far as shoulder height is concerned, there is no significant difference.

A comparison of these two results indicates that shoulder height does not increase significantly. Naturally it is the neck length which increases.

SIGNIFICANCE OF THE DIFFERENCES OF MEANS

PROBLEM 18.20. *How can "t" test be applied for testing the significance of the difference between two sample means. Calculate the value of t in the case of two characters A and B whose corresponding values are given below :—*

A	*41,*	*49,*	*34,*	*36,*	*49,*	*50,*	*36,*	*20,*	*18*
B	*46,*	*44,*	*30,*	*55,*	*26,*	*28,*	*29,*		

SOLUTION. Calculation of the mean and the standard deviation of the samples A and B

A			B		
Size of the item (X_1)	Deviation from the av. (37) (X_1-X_1)	Square of deviation $(X_1-\overline{X_1})^2$	Size of the item (X_2)	Deviation from the av. (34) $(X_2-\overline{X_2})$	Square of deviation $(X_2-\overline{X_2})^2$
41	+4	16	46	+12	144
49	+12	144	44	+10	100
34	−3	9	30	−4	16
36	−1	1	35	+1	1
49	+12	144	26	−8	64
50	+13	169	28	−6	36
36	−1	1	29	−5	25
20	−17	289			
18	−19	361			
$\Sigma X_1=333$		$\Sigma(X_1-\overline{X_1})^2=1134$	$\Sigma X_2=238$		$\Sigma(X_2-\overline{X_2})^2=386$

Mean of the sample A

$$\overline{X_1}=\frac{\Sigma X_1}{n_1}=\frac{333}{9}=37.$$

Mean of the sample B

$$\overline{X_2}=\frac{\Sigma X_2}{n_2}=\frac{238}{7}=34.$$

Hence the difference of the two sample means is

$$\overline{X}_1-\overline{X}_2=37-34=3.$$

The standard error of the difference is

$$S=\sqrt{\frac{\Sigma(X_1-\overline{X}_1)^2+(X_2-\overline{X}_2)^2}{n_1+n_2-2}}=\sqrt{\frac{1134+386}{9+7-2}}$$

$$=\sqrt{108{\cdot}57}=10{\cdot}4$$

Now, substituting the values in the formula

$$t=\frac{\overline{X}_1-\overline{X}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{37-34}{10{\cdot}4\sqrt{\frac{1}{9}+\frac{1}{7}}}={\cdot}577.$$

Thus, the value of t for given data is .577.

PROBLEM 18·21 *The figures below are for protein tests of the same variety of wheat grown in two districts. In district 1 the average for 5 samples is 12.74, and in district 2, the average for 7 samples is 13·03. If these are the only figures available, test the significance of the difference between the average proteins for the two districts.*

	Protein	*Results*					
District 1.................	*12·6*	*13.4*	*11·9*	*12·8*	*13·0*		
District 2.................	*13.1*	*13·4*	*12·8*	*13·5*	*13·3*	*12·7*	*12·4*

SOLUTION Calculation of the standard error of the difference of the average proteins for the two districts 1 and 2.

District 1			District 2		
Protein Results (X_1)	Deviation from the average (12·74) $(X_1-\overline{X}_1)$	Square of the deviation $(X_1-\overline{X}_1)^2$	Protein results (X_2)	Deviation from the average (14·03) $(X_2-\overline{X}_2)$	Square of the deviation $(X_2-\overline{X}_2)^2$
12·6	—.14	.0196	13·1	+·07	·0049
13·4	+·66	.4356	13.4	+.37	.1369
11.9	—.84	·7056	12.8	—·23	·0529
12·8	+·06	·0036	13·5	+·47	·2209
13·0	+·26	·0676	13.3	+.27	·0729
			12·7	—·33	·1089
			12·4	—·63	·3969
		$\Sigma(X_1-\overline{X}_2)^2$ =1.2320			$\Sigma(X_2-\overline{X}_2)$ =.9943

The difference between the average proteins for the two districts is 13.03 —12.74, or .29. The standard error of the difference is

$$S=\sqrt{\frac{\Sigma(x_1-x_1)^2+(x_2-x_2)^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{1.2320 + .9943}{5+7-2}} = \sqrt{.22263} = .472.$$

Substituting the values in the formula

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

We get,

$$t = \frac{12.74 - 13.03}{.472 \times \sqrt{\frac{1}{5} + \frac{1}{7}}} = -1.05$$

The number of degrees of freedom is $n_1 + n_2 - 2 = 5 + 7 - 2 = 10$.

The value of t for 10 degrees of freedom at 5% level of significance is 2.228. The calculated value is considerably less than this figure. Hence, it can be concluded that the average protein contained by wheat in the two districts does not differ significantly.

Note : ± signs have no significance while we compare the calculated and the table values of t.

PROBLEM 8.22. *Is the difference between the mean scores of the two groups, with 8 and 7 members respectively, significant from the following data :—*

1st group (8 students) x_1 : *19 14 13 16 19 18 16 17*
2nd group (7 students) x_2 : *21 19 16 22 18 20 19*

SOLUTION. Calculation of Mean and Variance

x_1	x_1^2	x_2	x_2^2
19	361	21	441
14	196	19	361
13	169	16	256
16	256	22	484
19	361	13	324
18	324	20	400
16	256	19	361
17	289		
132	2212	135	2627

$$\overline{x_1} = \frac{132}{8} = 16.5; \quad \overline{x_2} = \frac{135}{7} = 19.386$$

$$\Sigma(x_1 - \overline{x_1})^2 = 2212 - \frac{(132)^2}{8}$$

$$= 2212 - 2178 = 34$$

$$\Sigma\left(x_2 - \overline{x_2}\right)^2 = 2627 - \frac{(135)^2}{7}$$

$$= 2627 - 2603.57 = 23.43$$

$$\sigma^2 = \frac{\Sigma(x_1 - \overline{x}_1)^2 + \Sigma(x_2 - \overline{x}_2)^2}{n_1 + n_2 - 2}$$

$$= \frac{34 + 23.43}{8 + 7 - 2} = \frac{57.43}{13} = 4.42 \quad \therefore \sigma = 2.1$$

$$\therefore t = \frac{\overline{x_1} - \overline{x_2}}{S \times \sqrt{\frac{1}{n_1} \times \frac{1}{n_2}}} \text{ or } \frac{\bar{x}_1 - \bar{x}_2}{S} \sqrt{\frac{n_1 \times n_2}{n_1 + n_2}}$$

$$= \frac{16.5 - 19.286}{2.1} \sqrt{\frac{56}{15}} = \frac{-2.785}{2.1} \times 1.932$$

$$= 2.785 \times 0.92 = 2.56220 \text{ or } 2.56$$

Degrees of freedom $= 8 + 7 - 2 = 13$.

For 13 degrees of freedom, $t = 2.16$ is significant at .05 level. The observed value of t is greater than this and, therefore, the difference between the means is significant.

PROBLEM 18.23. *The heights of six randomly chosen sailors are in cms. 163, 165, 168, 169, 171 and 172. Those of ten randomly chosen soldiers are : 161, 162, 165, 166, 169, 169, 170, 171, 172 and 173. Discuss in the light that these data throw on the suggestion that soldiers are, on the average, taller than sailors.*

SOLUTION. Calculation of the standard error of the difference of the mean heights of the sailor and the soldier.

Sailors			Soldiers		
Height in centimetres (x_1)	Deviation from av. height (168 cms.) $(x_1-\overline{x_1})^2$	Square of deviation $(x_1-\overline{x_1})^2$	Height in centimetres (x_2)	Deviation from av. height (167.8 cms. $(x_2-\overline{x_2})$	Square of deviation $(x_2-\overline{x_2})^2$
163	-5	25	161	-6.8	46.24
165	-3	9	162	-5.8	33.64
168	0	0	165	-2.8	7.84
169	$+1$	1	166	-1.8	3.24
171	$+3$	9	169	$+1.2$	1.44
172	$+4$	16	169	$+1.2$	1.44
			170	$+2.2$	4.84
			171	$+3.2$	10.24
			172	$+4.2$	17.64
			173	$+5.2$	27.04
$\Sigma(x_1)=1008$	$\Sigma(x_1-\overline{x_1})=0$	$\Sigma(x_1-\overline{x_1})^2=60$	$\Sigma x_2=1678$	$\Sigma(x_2-\overline{x_2})=0$	$\Sigma(x_2-x_2)^2=153.60$

Mean height of the sailor

$$\overline{x_1}=\frac{\Sigma x_1}{n_1}=\frac{1008}{6}=168 \text{ centimeters}$$

Mean height of the soldier

$$\overline{x_2}=\frac{\Sigma x_2}{n_2}=\frac{1678}{10}=167.8 \text{ centimetres}$$

Difference between the two mean heights is

$\overline{x_1}-\overline{x_2}=168-167.8$, or .2 cms.

The standard error of this difference is

$$S=\sqrt{\frac{\Sigma(x_1-\overline{x}_1)^2+\Sigma(x_2-\overline{x_2})^2}{n_1+n_2-2}}=\sqrt{\frac{60+153.6}{6+10-2}}=15.26=3.9$$

Substituting these values in the formula

$$t=\frac{\overline{x_1}-\overline{x_2}}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$=\frac{68-67.8}{3.9\sqrt{\frac{1}{6}+\frac{1}{10}}}=.099$$

The number of degrees of freedom is

$n_1+n_2-2=6+10-2$, or 14.

The value of t for 14 degrees of freedom at 5% level of significance is 2.145. The calculated value is much less than this value. So the difference between the mean heights of sailors and soldiers is not significant. Hence the suggestion that soldiers are on the average taller than sailors is wrong.

PROBLEM 18.24. *Two horses, A and B, tested according to the times (in seconds) taken to run a particular track with the following results :*

Horse A	*28,*	*30,*	*32,*	*33,*	*33,*	*29, and*	*34*
Horse B	*29,*	*30,*	*30.*	*24.*	*27,*	*and 29.*	

Analyse the above data and report whether or not you can discriminate between the two horses.

SOLUTION. Calculation of the standard error of the difference in records of Horse A and B.

Horse A			Horse B		
Records (x_1)	Deviation from mean (31.3) $(x_1-\overline{x}_1)$	Square of deviation $(x_1-\overline{x}_1)^2$	Records (x_2)	Deviation from mean (28.2) $(x_2-\overline{x}_2)$	Square of deviation $(x_2-\overline{x}_2)^2$
28	-3.3	10.89	29	$+.8$	.64
30	-1.3	1.69	30	$+1.8$	3.24
32	$+.7$	.49	30	$+1.8$	3.24
33	$+1.7$	2.89	24	$+4.2$	17.64
33	$+1.7$	2.89	27	-1.2	1.44
29	-2.3	5.29	29	$+.8$	.64
34	$+2.7$	7.29			
$\Sigma x_1=219$		$\Sigma(x_1-\overline{x}_1)^2 =31.43$	$\Sigma x_2=169$		$\Sigma(\overline{x_2-x_2}) =-26.48$

Mean record of the horse A,

$$x_1=\frac{\Sigma(x_1)}{n_1}=\frac{219}{7}=31.3$$

Mean record of the horse B.

$$x_2=\frac{\Sigma(x_2)}{n_2}=\frac{169}{6}=28.2$$

The difference in two means $=(31.3-28.2)=3.1$

The standard error of this difference,

$$S=\sqrt{\frac{(x_1-\overline{x}_1)^2+(x_2-\overline{x}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{31.43+26.84}{7+6-2}}$$

$$=\sqrt{\frac{58.21}{11}}=\sqrt{5.297}=2.302$$

Substituting these values in the formula.

$$t=\frac{\overline{x}_1-\overline{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{3.1}{2.302\sqrt{\frac{1}{7}+\frac{1}{6}}}=2.42$$

The number of degrees of freedom is $n_1+n_2-2=7+6=2=11$.

The value of t for 11 degrees of freedom at 5% level of probability=2·20. The calculated value is more than this, hence the difference is significant. *i.e.* there is difference between the two horses.

PROBLEM 18.25. *In a test examination given to two groups of students, the marks obtained were as follows :*

First Group — *18, 20, 36, 50, 49, 36, 34, 49, 41.*
Second Group— *29, 28, 26, 35, 30, 44, 46.*

Examine the significance of difference between the arithmetic average of the marks secured by the students of the above two groups.

SOLUTION. Calculation of the mean and the standard deviations of the two Groups.

First Group			Second Group		
Size of the item (x_1)	Deviation from the av. 37 $(x_1-\overline{x}_1)$	Square of deviation $(x_1-\overline{x}_1)^2$	Size of item (x_2)	Deviation from the av. 34 $(x_2-\overline{x}_2)$	Square of deviation $(x_2-\overline{x}_2)^2$
18	−19	361	29	−5	25
20	−17	289	28	−6	36
36	−1	1	26	−8	64
50	+13	169	35	+1	1
49	+12	144	30	−4	16
36	−1	1	44	+10	100
34	−3	9	46	+12	144
49	+12	144			
41	+4	16			
$\Sigma x_1=333$		$\Sigma(x_1-\overline{x})^2=1134$	$\Sigma x_2=238$		$(x_2-\overline{x}_2)^2=386$

Mean of the First Group

$$\overline{x}_1=\frac{\Sigma x_1}{n_1}=\frac{333}{9}=37$$

Mean of the Second Group

$$\overline{x}_2=\frac{\Sigma x_2}{n_2}=\frac{238}{7}=34$$

Hence the difference of the two sample means is

$\overline{x}_1-\overline{x}_2=37-34=3.$

The standard error of the difference is

$$S=\sqrt{\frac{\Sigma(x_1-\overline{x}_1)^2+\Sigma(x_2-\overline{x}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{1134+386}{9+7-2}}=\sqrt{108.57}=10.42$$

After substituting these values in the formula

$$t=\frac{\overline{x}_1-\overline{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{37-34}{10.42\sqrt{\frac{1}{9}+\frac{1}{7}}}=.577$$

The value of t for the given data is .577.

The degrees of freedom are n_1+n_2-2 or $9+7-2$ or 14. The value of t for 14 degrees of freedom at 5% level of significance is 2.14 which is very much more than the observed value. The difference is, therefore, insignificant.

PROBLEM 18.26. *What is 't' test for small samples ? Explain, giving the suitable formula, how this test can be applied for testing the significance of the difference between two sample means. Hence calculate the value of 't' in the case of two characters A and B whose values are given below :*

A— 16, 10, 8, 9, 9, 8.
B— 8, 4, 5, 9, 12, 4.

SOLUTION. Calculation of the mean and the standard deviation of the samples A and B.

A			B		
Size of item (x_1)	Deviation from the av. (10) $(x_1-\overline{x}_1)$	Square of deviation $(x_1-\overline{x}_1)^2$	Size of item (x_2)	Deviation from the av. (7) $(x_2-\overline{x}_2)$	Square of deviation $(x_2-\overline{x}_2)^2$
16	+6	36	8	+1	1
10	0	0	4	−3	9
8	−2	4	5	−2	4
9	−1	1	9	+2	4
9	−1	1	12	+5	25
8	−2	4	4	−3	9
$\Sigma x_1=60$		$\Sigma(x_1-\overline{x}_1)^2=46$	$\Sigma x_2=42$	$\Sigma x_2=42$	$\Sigma(x_2-\overline{x}_2)^2=52$

Mean of sample A.

$$\overline{x}_1=\frac{\Sigma x_1}{n_1}=\frac{60}{6}=10$$

Mean of sample B.

$$\overline{x}_2=\frac{\Sigma x_2}{n_2}=\frac{42}{6}=7$$

Hence the difference of the two sample means is

$$(\overline{x}_1-\overline{x}_2)=10-7=3$$

The standard error of the difference is

$$S=\sqrt{\frac{\Sigma(x_1-\overline{x}_1)^2+\Sigma(x_2-\overline{x}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{46+52}{6+6-2}}=\sqrt{\frac{98}{10}}=\sqrt{9.8}=3.13$$

Now, substituting these values in the formula.

$$t=\frac{\overline{x}_1-\overline{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{10-7}{3.13\sqrt{\frac{1}{6}+\frac{1}{6}}}=1.65$$

The value of t for given data is 1.65. The degrees of freedom in the question are 10. The 5% value of t for 10 degrees of freedom is 2.23. The calculated value is less than this and so the difference between the two means is not significant.

PROBLEM 18.27. *For a random sample of 10 pigs, fed on diet A, the creases in weight in kilograms in a certain period were :*

10, 6, 16, 17, 13, 12, 8, 14, 15, 9.

For another random sample of 12 pigs, fed on diet B, the increases in the same period were :

7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17.

Test whether diets A and B differ significantly as regards their effects on increases in weight. You may use the following extract from statistical table —

Degrees of freedom	*Value of 't' significant at 5% level of probability*
19	2.09
20	2.09
21	2.08
22	2.07
23	2.07

SOLUTION. Calculation of the standard error of the difference in mean gains in weight on diet A and diet B.

Diet A			Diet B		
Gains in weight (x_1)	Deviations from mean (12) ($x_1-\bar{x}_1$)	$(x_1-\bar{x}_1)^2$	Gains in weight (x_2)	Deviations from mean (15) ($x_2-\bar{x}_2$)	$(x_2-\bar{x}_2)^2$
10	−2	4	7	−8	64
6	−6	36	13	−2	4
16	+4	16	22	+7	49
17	+5	25	15	0	0
13	+1	1	12	−3	9
12	0	0	14	−1	1
8	−4	16	18	+3	9
14	+2	4	8	−7	49
15	+3	9	21	+6	36
9	−3	9	23	+8	64
			10	−5	25
			17	+2	4
$\Sigma x_1=120$		$\Sigma(x_1-\bar{x}_1)^2=120$	$\Sigma x_2=180$		$\Sigma(x_2-\bar{x}_2)^2=314$

Mean gain in weight on diet A.

$$\bar{x}_1=\frac{\Sigma(x_1)}{n_1}=\frac{120}{10}=12 \text{ kilograms}$$

Mean gain in weight on diet B

$$\bar{x}_2=\frac{\Sigma(x_2)}{n_2}=\frac{180}{12}=15 \text{ kilograms}$$

The difference in two means=(15—12)=3 kilograms

The standard error of this difference

$$S=\sqrt{\frac{(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_1-2}}$$

$$=\sqrt{\frac{120+314}{10+12-2}}=\sqrt{\frac{434}{20}}$$

$$21.7=4.65$$

Substituting these values in the formula.

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{12-15}{4.65\sqrt{\frac{1}{10}+\frac{1}{12}}}=1.5.$$

The number of degrees of freedom is $n_1+n_2-2=10+12-2=20$. The value of t for 20 degrees of freedom at 5% level of probability=2.09. The calculated value is less than this and hence the difference is not significant.

PROBLEM 18.28. *Eight pots growing three wheat plants each were exposed to a high tension discharge, while nine similar pots were enclosed in an earthen ware case. The number of tillers in each pot were as follows :*

Caged	*17*	*26*	*18*	*25*	*27*	*28*	*26*	*23*	*17*
Electrified	*16*	*16*	*22*	*16*	*21*	*18*	*15*	*20*	

See whether electrification exercises any real effect on the tillering by using t test of significance

SOLUTION. Calculation of the standard error of the difference of the average tillers of the two samples.

Caged				Electrified			
Pots	Tillers in each pot x_1	Deviation from the average tiller (23) $(x_1-\bar{x}_1)$	Square of deviation $(x_1-\bar{x}_1)^2$	Pots	Tillers in each pot x_2	Deviation from the average tiller (18) $(x_2-\bar{x}_2)$	Square of deviation $(x_2-\bar{x}_2)^2$
1	17	−6	36	1	16	−2	4
2	26	+3	9	2	16	−2	4
3	18	−5	25	3	22	+4	16
4	25	+2	4	4	16	−2	4
5	27	+4	16	5	21	+3	9
6	28	+5	25	6	18	+0	0
7	26	+3	9	7	15	−3	9
8	23	0	0	8	20	+2	4
9	17	−6	36				
$n_1=9$	$\Sigma(x_1)$ =207		$\Sigma(x_1-\bar{x}_1)^2$	$n_2=8$	$\Sigma(x_2)$ = 144		$\Sigma(x_2-\bar{x}_2)^2$ =50

Mean tillers of caged sample

$$\bar{x}_1=\frac{\Sigma(x_1)}{n_1}=\frac{207}{9}=23.$$

Mean tillers of 'Electrified' sample

$$\bar{x}_2=\frac{\Sigma(x_2)}{n_2}=\frac{144}{8}=18.$$

Difference between the two mean tillers is

$$\bar{x}_1-x_2=23-18=5.$$

The standard error of this difference is

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{160+50}{9+8-2}}=\sqrt{14}=3.74.$$

Substituting these values in the formula

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

e get,

$$t=\frac{23-18}{3.74\times\sqrt{\frac{1}{9}+\frac{1}{8}}}=2.751$$

The number of degrees of freedom is

$$n_1+n_2-2=9+8-2=15.$$

The value of t for 15 degrees of freedom at 5% level of significance is 131 which is less than the calculated value of t. Hence, the difference between e mean tillers of the two samples is significant which is indicative of the fact at electrification does exert some effect on the tillering.

PROBLEM 18.29. *The ash content of coal from two different mines was alysed, five analysis being made of the coal from the first mine, four of that from e second mine. Are we justified in supposing that the two mines consist coal ith the same percentage of ash content on the basis of the results obtained, which e recorded in the following tables.*

Table A *per cent. ash content*	*Table B* *per cent. ash content*
24.3	*18 2*
20.8	*16.9*
23.7	*20.2*
21.3	*16.7*
17.4	

SOLUTION. Calculation of the standard error of the mean percentages of sh content in the two mines.

First Mine				Second Mine			
No. of analysis	Percentage of ash content (x_1)	Deviation from the av. percentage $(x_1-\bar{x}_1)$	Square of deviation $(x_1-\bar{x}_1)^2$	No. of analysis	Percentage of ash content (x_2)	Deviation from av. percentage $(x_2-\bar{x}_2)$	Square of deviation $(x_2-\bar{x}_2)^2$
1	24.3	+2.8	7.84	1	18.2	+0.2	.04
2	20.8	—.7	.49	2	16.9	—1.1	1.21
3	2.7	+2.2	4.84	3	20.2	+2.2	4.84
4	21.3	—.2	.04	4	16.7	—1.3	1.69
5	17.4	—4.1	16.81				
$n_1=5$	$\Sigma(x_1)=107.5$		$\Sigma(x_1-\bar{x}_1)^2=30\ 02$	$n_2=4$	$\Sigma(x_2)=72.0$		$\Sigma(x_2-\bar{x}_2)^2=7.78$

Mean percentage of ash content of coal of the First Mine

$$\bar{x}=\frac{\Sigma(x_1)}{n_1}=\frac{107.5}{5}=2.15$$

Mean percentage of ash content of coal of the Second Mine

$$\bar{x}_2=\frac{\Sigma(x_2)}{n_2}=\frac{72.0}{4}=18.$$

The difference in mean percentages of ash content of coal in First and Second Mine is $\bar{x}_1-\bar{x}_2=21.5=18=3.5$

The standard error of this mean difference is

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{30.02+7.78}{5+4-2}}=\sqrt{5.4}=2.324.$$

Substituting these values in the formula,

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$=\frac{21.5-18}{2.324\sqrt{\frac{1}{5}+\frac{1}{4}}}=2.248.$$

The number of degrees of freedom is

$$n_1+n_2-2=5+4-2=7.$$

The value of t for 7 degrees of freedom at 5% level of significance is 2.365. The calculated value is less than this value, so the difference in the percentage of ash content of coal of the two mines is not significant. Hence, the inference may be drawn that the two mines consist of coal with the same percentage of ash content.

PROBLEM 19.30. *Is there a significant difference in the strength of lead in two number 2 Pencils manufactured by R. G. T. Guteba & Co.*

Test	*Pencil (a) Strength in kilograms*	*Test*	*Pencil (b) Strength in kilograms*
1	*1.62*	*1*	*1.78*
2	*1.72*	*2*	*1.48*
3	*1.68*	*3*	*1.72*
4	*1.50*	*4*	*1.62*
Total	*6.52*	*Total*	*6.60*

SOLUTION. Calculation of the standard error of the difference in the mean strength of the two pencils A and B.

Pencil A				Pencil B			
No. of tests	Strength in Kilograms	Deviation from the av.(1.63)	Square of deviation	No. of tests	Strength in Kilograms	Deviation from the av (1.65)	Square of deviation
	(x_1)	$(x_1-\bar{x}_1)$	$(x_1-\bar{x}_1)^2$		(x_2)	$(x_2-\bar{x}_2)$	$(x_2-\bar{x}_2)^2$
1	1.62	—.01	.0001	1	1.78	+.13	.0169
2	1.72	+.09	.0011	2	1.48	—.17	.0289
3	1.68	+.05	.0025	3	1.72	+.07	.0049
4	1.50	—.13	.0169	4	1.62	— 03	.0009
$n_1=4$	(Σx_1) $=6.52$		$\Sigma(x_1-\bar{x}_1)^2$ $=0.0276$	$n_2=4$	$\Sigma(x_2)=$ 6.60		$\Sigma(x_2-\bar{x}_2)^2$ $=.0516$

Mean strength of lead in Pencil A is

$$\bar{x}_1=\frac{\Sigma(x_1)}{n_1}=\frac{6.52}{4}=1.63 \text{ Kilograms}$$

Mean strength of lead in Pencil B is

$$\bar{x}_2=\frac{\Sigma(x_2)}{n_2}=\frac{6.60}{4}=1.65 \text{ Kilograms}$$

The difference in the mean strength of the two pencils is

$$\bar{x}_1-\bar{x}_2=1.63-1.65=.02 \text{ Kilograms.}$$

The standard error of this difference is

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

$$\sqrt{\frac{0.276+.0516}{4+4-2}}$$

$$\sqrt{.0132}=.115$$

Substituting these values in the formula

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n}}}$$

We get,

$$t=\frac{1{\cdot}63-1.65}{{\cdot}115\sqrt{\frac{1}{4}+\frac{1}{4}}}=\frac{-{\cdot}02}{{\cdot}0805}=-.248$$

The number of degrees or freedom is $n_1+n_2-2=4+4-2$, or 6.

The value of t for 6 degrees of freedom at 5% level of significance is 2.447. The calculated value of t is(slightly greater than this value. Hence the difference in the strength of lead in the two types of pencils is significant.

Note—In the calculated value of t ± signs should be ignored.

PROBLEM 18.31. *Below are given the gains in weight (kilograms) of hogs fed on two different diets. Twelve animals were fed on diet A, 15 on diet B. Is either diet superior ?*

Gains in weight on diet A : 25, 32, 30, 34, 24, 25, 14, 32, 24, 30, 31, 35.

Gains in weight on diet B : 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22.

SOLUTION. Calculation of the standard error of the difference in the mean gains in weight on diet A and diet B.

	Diet A				Diet B		
No. of Hogs	Gains in weight in kilograms (X_1)	Deviation from the av. (28) ($X_1-\bar{X}_1$)	Square of deviation ($X_1-\bar{X}_1)^2$	No. of Hogs	Gains in weight in kilograms (X_2)	Deviation from the av. ($X_2-\bar{X}_2$)	Square of deviation $(X-\bar{X}_2)^2$
1	25	−3	9	1	44	+14	196
2	32	+4	16	2	34	+4	16
3	30	+2	4	3	22	−8	64
4	34	+6	36	4	10	−20	400
5	24	−4	16	5	47	+17	289
6	25	−3	9	6	31	+1	1
7	14	−14	196	7	40	+10	100
8	32	+4	16	8	30	0	0
9	24	−4	16	9	32	+2	4
10	30	+2	4	10	35	+5	25
11	31	+3	9	11	18	−12	144
12	35	+7	49	12	21	−9	81
				13	35	+5	25
				14	29	−1	1
				15	22	−8	64
$n_1=12$	$\Sigma(X_1)=336$		$\Sigma(X_1-\bar{X}_1)^2=380$	$n_2=15$	$\Sigma(X_2)=450$		$\Sigma(X_2-\bar{X}_2)=1410$

Mean gain in weight on diet A

$$\bar{X}_1 = \frac{\Sigma(X_1)}{n_1} = \frac{336}{12} = 28 \text{ kilograms}$$

Mean gain in weight on diet B

$$\bar{X}_2 = \frac{\Sigma(X_2)}{n_2} = \frac{450}{15} = 30 \text{ kilograms}$$

The difference in the mean gain in weight on two different diet is $\bar{X}_1—\bar{X}_2=$ 28—30 = 2 kilograms

The standard error of this difference is

$$= \sqrt{\frac{\Sigma(X_1—\bar{X}_1)^2 + \Sigma(X_2—X_2)^2}{n_1+n_2—2}}$$

$$= \sqrt{\frac{380+1410}{12+15—2}} = \sqrt{71.6} = 8{\cdot}46.$$

Substituting these values in the formula

$$t = \frac{\bar{X}_1—\bar{X}_2}{S \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

We get

$$t = \frac{—2}{8.46\sqrt{\frac{1}{12}+\frac{1}{15}}} = {\cdot}611.$$

The number of degrees of freedom is $n_1+n_2—2=12+15—2$, or 25. The value of t for 25 degrees of freedom at 5% level of significance is 2.06. The calculated value of t is considerably less than this value. Hence the difference in the mean gain in weight on diets A and B is insignificant.

PROBLEM 18.32. *What is t test in small sample theory ? How does Student's t differ from Fisher's t ?*

A group of seven week old chickens, reared on a high protein diet, weigh 12, 15, 11, 16, 14, 14, 16 ounces; a second group of 5 chickens, similarly treated except that they receive a low protein diet, weigh 8, 10, 14, 10, 13 ounces. Calculate the value of t and test whether there is significant evidence that additional protein has increased the weight of the chickens.

Value of t on levels of significance 'P'

Degree of freedom	*P=0.23*	*P=0.81*
10	*1.23*	*1.81*
11	*2.20*	*1.80*
12	*2.18*	*1.78*

SOLUTION. Calculation of the mean and standard deviation of the two groups of chickens.

	Group A				Group B		
Chicken	Weight (x_1)	Deviation from the av. (14) $(x_1-\bar{x}_1)$	Square of deviation $(x_1-\bar{x}_1)^2$	Chicken	Weight (x_2)	Deviation from the av. (11) $(x_2-\bar{x}_2)$	Square of deviation $(x_2-\bar{x}_2)^2$
1	12	-2	4	1	8	-3	9
2	15	$+1$	1	2	10	-1	9
3	11	-3	9	3	14	$+3$	9
4	16	$+2$	4	4	10	-1	1
5	14	0	0	5	13	$+2$	4
6	14	0	0				
7	16	$+2$	4				
$n_1=7$	$\Sigma(x_1)=98$		$\Sigma(x_1-\bar{x}_1)^2=22$	$n_2=5$	$\Sigma x_2=55$		$\Sigma(x_2-\bar{x}_2)^2=24$

Mean of the Group A

$$\bar{x}_1=\frac{\Sigma x_1}{n_1}=\frac{98}{7}=14$$

Mean of the Group B

$$\bar{x}_2=\frac{\Sigma x_2}{n_2}=\frac{55}{5}=11$$

Hence the difference of the two sample means is

$$x_1-\bar{x}_2=14-11=3$$

The standard error of the difference is

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_2)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{22+24}{7+5-2}}\qquad=\sqrt{\frac{46}{10}}$$

$$=\sqrt{4.6}=2.14$$

Now, substituting these values in the formula

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{14-11}{2.14\sqrt{\frac{1}{7}+\frac{1}{5}}}=2.42$$

The value of t for the given data is 2.39.

The number of degrees of freedom is

$$n_1+n_2-2=7+5-2=10$$

The value of t for 10 degrees of freedom at 5% level of significance is 2.23 The calculated value of t is more than the table value and hence the difference between the two means is significant and so it can be reasonably concluded that additional protein has increased the weight of the chicken.

PROBLEM 18.33. *The means of two random samples of sides 9 and 7 respectively are 196.42 and 198.82 respectively. The sums of the square of the deviation from the mean are 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population ?*

d. f	*5% value of t*	*1% value of t*
13	*2.160*	*3.012*
14	*2.145*	*2.977*
15	*2.131*	*2.947*
16	*2.120*	*2.921*

SOLUTION. The standard error of the difference of the means of the two samples is

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

Substituting the values, we get,

$$S=\sqrt{\frac{26.94+18.73}{9+7-2}}=\sqrt{\frac{45.67}{14}}=\sqrt{3.26}=1.8$$

Applying to the given data, the formula

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{196.42-198.82}{1.8\times\sqrt{\frac{1}{9}+\frac{1}{7}}}=\frac{-2.5}{.9}=2.667$$

The number of degrees of freedom is $n_1+n_2-2=9+7-2=14$. The value of t for 14 degrees of freedom at 5% level or significance is 2.145. This value is less than the calculated value of t which is 2.667. Hence the difference between the two means is significant and the samples cannot be considered to have been drawn from the same normal population.

PROBLEM 18 34. *In three samples of 50 lines each from Shakespeare's "Romeo and Juliet" (an early play), the following numbers of weak endings were observed : 7, 9 and 10. In three similar samples from "Cymbeline" (late), the numbers of weak endings were 15, 11, 12. Discuss the suggestion that Shakespeare's prosody, as judged by the number of weak endings, changed with advancing years.*

SOLUTION. Calculation of the standard error of the difference of the mean weak endings of two plays of Shakespeare.

Romeo and Juliet (an early play)				Cymbeline (a late play)			
No. of sample	Number of weak endings x_1	Deviation from the av.(8.67) $(x_1-\bar{x}_1)$	Square of deviation $(x_1-\bar{x}_1)^2$	No. of sample	Number of weak endings (x_2)	Deviation from the average (12.67) $(x_2-\bar{x}_2)$	Square of deviation $(x_2-\bar{x}_2)^2$
1	7	−1.67	2.7889	1	15	+2.33	5.4289
2	9	+ .33	.1089	2	11	−1.66	2.7556
3	10	+1 33	1.7689	3	12	—·67	.4489
$n_1=3$	$\Sigma(x_1)$ =26		$\Sigma(x_1-\bar{x}_1)^2$ =4.6667	$n_2=3$	$\Sigma(x_2)$ = 38		$\Sigma(x_2-\bar{x}_2)^2$ =8.6334

Mean weak endings of Romeo and Juliet, the early play

$$\bar{x}_1=\frac{\Sigma(x_1)}{n_1}=\frac{26}{3}=8.67$$

Mean weak endings of Cymbeline, late play

$$\bar{x}_2=\frac{\Sigma(x_2)}{n_2}=\frac{38}{3}=12.67$$

Thus, the difference between the two mean weak endings is $\bar{x}_1-\bar{x}_2=8.67-12.67$, or -4,

The standard error of this difference is

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{4.6667+8.6334}{3+3-2}}=\sqrt{3.325}=1.823.$$

Substituting these values in the formula,

$$t=\frac{\bar{x}_1-\bar{x}_2}{S=\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{-4}{1.823\times\sqrt{\frac{1}{3}+\frac{1}{3}}}=-2.688$$

The number of degrees of freedom is $n_1+n_2-2=3+3-2$ or 4

The value of t for 4 degrees of freedom at 5% level of significance is 2.776 and at 10% level of significance is 4.13. The calculated value of t is less than the table value and hence the difference between two means is not significant and it cannot be reasonably concluded that Shakespeare's prosody, as judged by the number of weak endings, changed with advancing years.

Note. ±signs have no value in comparing calculated and table values of t.

PROBLEM 18.35. *Mitchell conducted a paired feeding experiment with pigs on the relative value of limestone and bonemeal for bone development. The results are given below.*

Ash content in Percentage of Scapulas of Pairs of Pigs fed on Limestone and Bonemeal.

Pair	*Limestone*	*Bonemeal*
1	*49·2*	*51.5*
2	*53.3*	*54.9*
3	*50.6*	*52.2*
4	*52.0*	*53.3*
5	*46.8*	*51.6*
6	*50.5*	*54.1*
7	*52.1*	*54.2*
8	*53.0*	*54.3*
Mean	*50.94*	*53.14*

Determine the significance of the difference between the means in two ways : (1) by assuming that the values are paired, and (2) by assuming that the values are not paired. On the basis of your results, discuss the effect of pairing.

SOLUTION.

(a) *When values are paired.*

Calculation of the mean and the standard deviations of the differences of ash content in the two samples.

Pair	Ash content in percentage of scapulas of pairs of pigs		Difference (Bonemeal-Limestone)	Deviation from the average (2.3)	Square of deviation
	Fed on Limestone	Fed on Bonemeal	(m)	(d)	(d^2)
1	49.2	51.5	2.3	+ 0	0
2	53.3	54.9	1.6	— .7	.49
3	50.6	52.2	1.6	— .7	.49
4	52.0	53.3	1.3	—1.0	1.00
5	46·8	51.6	4.8	+2.5	6.25
6	50.5	54.1	3.6	+1.3	1.69
7	52.1	54.2	2.1	— .2	.04
8	53.0	54.3	1.3	—1.0	1.00
			$\Sigma m=18.6$		$\Sigma d^2=10.96$

Mean of the differences of ash content in two samples

$$\bar{x}=\frac{\Sigma m}{x}=\frac{18.6}{8}=2.3 \text{ approx.}$$

Standard deviation of the differences in the ash content of two samples

$$\sigma=\sqrt{\frac{\Sigma d^2}{n-1}}=\sqrt{\frac{10.96}{7}}=1.25$$

Now, we take as our hypothesis, the supposition that the ash contents in percentage of scapulas of pairs of pigs fed on limestone and bonemeal are not different, that is, the mean of the differences of the population or $m=0$.

Substituting these values in the formula,

$$t=\frac{\bar{x}-m}{\sigma}\sqrt{n}$$

We get,

$$t=\frac{2.3-0}{1.25}\sqrt{8}=5.2$$

The number of degrees of freedom is 8—1 or 7.

The value of t for 7 degrees of freedom at 5% level of significance is 2.365, which is considerably smaller than the calculated value. Hence the hypothesis is wrong and the difference between the two means is significant.

(*b*) *When values are not paired,*

Calculation of the standard error of the difference between the two means.

No. of pair	Ash content fed on lime-stone (x_1)	Deviation from the av.(50.94) $(x_1-\bar{x}_1)$	Square of deviation $(x_1-\bar{x}_1)^2$	No. of pair	Ash content fed on Bone-meal (x_2)	Deviation from the av.(53.14) $(x_2-\bar{x}_2)$	Square of deviation $(x_2-\bar{x}_2)^2$
1	49.2	—1.74	3.0276	1	51.5	—1.64	2.6896
2	53.3	+2.36	5.5696	2	54.9	+1.76	3.0976
3	50.6	—0.34	0.1156	3	52.2	—0.94	0.8836
4	52.0	+1.06	1.1236	4	53.3	+0.16	0.0256
5	46.8	—4.14	17.1396	5	51.6	—1.54	2.3716
6	50.5	—0.44	0.1936	6	54.1	+0.96	0.9216
7	52.1	+1.16	1.3456	7	54.2	+1.06	1.1236
8	53.0	+2.06	4.1436	8	54.3	+1.16	1.3456
$n_1=8$			$\Sigma(x_1-\bar{x}_1)^2=32.6588$	$n_2=8$			$\Sigma(x_2-\bar{x}_2)^2=12.4588$

Mean ash content in percentage of Scapulas of pairs of pigs fed on limestone is

$\bar{x}_1=50.94$

Mean ash content in percentage of Scapulas of pairs of pigs fed on bonemeal is

$\bar{x}_2=53.14.$

Thus the difference between the two percentages is $\bar{x}_1-\bar{x}_2=50.94-53.14$ $=2.20$

The standard error of this difference is

$$S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

Substituting the values we get,

$$S=\sqrt{\frac{32.6588+12.4588}{8+8-2}}=1{\cdot}79$$

Substituting these values in the formula

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{2.20}{1.79\sqrt{\frac{1}{8}+\frac{1}{8}}}=2.46$$

The number of degrees of freedom is

$n_1+n_2-2=8+8-2=14$

The value of t for 14 degrees of freedom at 5% level of significance is 2.1 4. The calculated value of t is more than the table value and hence the difference between the two means is significant and so it can be reasonably concluded that the relative value of limestone and bonemeal for bone development differs.

PROBLEM 18.36. *Two types of batteries are tested for their length of life and following results are obtained :—*

	No. in sample	*Mean*	*Variance*
Battery A	*10*	*500 hours*	*100*
Battery B	*10*	*560 hours*	*121*

Is there a significant difference in the two means ?

SOLUTION. On the hypothesis that the two samples are drawn from populations with identical means and variance, the value of combined standard deviation or

$$S \text{ or } \sigma p=\sqrt{\frac{n_1\sigma_1+n_2\sigma_2{}^2}{n_1+n_2-1}}$$

$$\sqrt{\frac{(10\times100)+20\times121)}{10+10-2}}=\sqrt{122.8}=11.08.$$

Substituting these values in the formula

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get,

$$t=\frac{506-560}{11.08\sqrt{\frac{1}{10}+\frac{1}{10}}}=12.2.$$

For 18 degrees of freedom at 5% level of significance the value of t is 2.1. The calculated value being much higher than this the difference is highly significant.

PROBLEM 18.37. *Two yarns spun to the same count are tested for their strength. The following results are obtained :—*

	No. in Sample	*Sample mean*	*Sample Standard Dev.*
Yarn A	*9*	*42*	*7.5*
Yarn B	*4*	*50*	*6.5*

Is the difference in mean strength significant :

SOLUTION. On the hypothesis that the two samples are drawn from populations which are identical both as to mean and standard deviation, the combined standard deviation would be

$$S \text{ or } \sigma p=\sqrt{\frac{n_1\sigma_1{}^2\times n_2\sigma_2{}^2}{n_1+n_2-2}}=\sqrt{\frac{9\times(7.5)^2+4(6.5)^2}{9+4-2}}=7.8$$

Substituting the values in the formula

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

We get

$$t=\frac{8}{4.67}=1.7$$

For 11 degrees of freedom (9+4—2) at 5% level of significance the value of $t=2{\cdot}2$. The calculated value being less than this the difference is not significant.

PROBLEM 18.38. *A set of 15 observation gives mean $=68.57$; standard deviation $=2{\cdot}40$; another of 7 observations gives mean $=64.14$; standard deviation $=2.70$.*

Use the t test to find whether the two sets of data were drawn from populations with the same mean, it being assumed that the standard deviations in the two populations were (a) equal, (b) not equal.

You may base your answer on the following extract from t table.

Significance limits of t probability of a value numerically greater than t.

Degrees of freedom	*.05*	*·02*	*·01*
6	*2.447*	*3.143*	*3.707*
14	*2.145*	*2.624*	*2.977*
20	*2.086*	*2.528*	*2.845*

SOLUTION. On the hypothesis that the two samples are drawn from the population which are identical both as to mean and standard deviation, the combined standard deviation, would be,

$$S \text{ or } \sigma=\sqrt{\frac{n_1\sigma_1{}^2+n_2\sigma_2{}^2}{n_1+n_2-2}}=\sqrt{\frac{15\times(2.40)^2+7(2.70)^2}{15+7-2}}$$

$$= \sqrt{\frac{(86.4+351.03)}{20}} = 2.62$$

Substituting these values in the formula.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

We get,

$$t = \frac{68.57 - 64.14}{2{\cdot}62\sqrt{\frac{1}{18}+\frac{1}{7}}} = \frac{4.43}{2.62 \times .458} = 3.7$$

For twenty degrees of freedom at 5% level of significance the value of t is 2086. The calculated value being higher than this the difference is significant.

PROBLEM 18.39. *A random sample of 10 boys had the following I.Q.*

70, 120, 110, 101, 88, 83, 95, 98, 107, 100
Another sample of 10 boys had the following I.Q.
65, 70, 75, 80, 92, 87. 85, 92, 93, 88

(*i*) *Do the first set of data support the assumption of a population mean I.Q. of 100 ?*

(*ii*) *Find the 95% fiducial limits of the population mean corresponding to the first set of data.*

(*iii*) *Can these be random samples from the same normal population ?*

SOLUTION. (*i*) We have to find out whether the difference between the sample mean and the population mean I.Q. of 100 is significant or not.

$$\text{Sample Mean} = \overline{X}_1 = \frac{70+120+110+101+88+83+95+98+107+100}{10}$$

$$= \frac{972}{10} = 97{\cdot}2$$

$\Sigma(X_1 - \overline{X}_1)^2$ or

$$= \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n} \;(4900+14400+12100+10201+7744+6889+9025+9604+11449+10000) - 94478.4 = 96312 - 94478.4 = 1833.6$$

$$\therefore \sigma^2 = \frac{1833.6}{9} \quad \therefore a = \frac{42.819}{3} = 14.273$$

$$t = \frac{(97.2-100)\sqrt{10}}{14.273} = \frac{2.8 \times 3.162}{14{\cdot}273} = \frac{8{\cdot}854}{14.273} = 0{\cdot}62$$

Degrees of freedom $= 10 - 1 = 9$

For 9 degrees of freedom $t = 0\,70$ is significant at 0.50. level of significance. The observed value of t is less than this and, therefore, not at all significant. Data do not throw any doubt on the assumption.

(*ii*) 95% fiducial limits of the Population Mean :

For 9 degrees of freedom, 95% or ·05 level value of $t = 2.26$.

∴ 95% fiducial limits of the population mean—

$$= 97.2 \pm \frac{14{\cdot}273}{\sqrt{10}} \times 2.26$$

$$= 97.2 \pm 10.2 \text{ or } 107.4 \text{ and } 87.0 \text{ Ans.}$$

(*iii*) We have already found that for the first sample $\bar{X}_1 = 97.2$ and $\Sigma(X_1 - \bar{X}_1)^2 = 1833.6$

For the second sample, we find :

$$\bar{X}^2 = \frac{65+70+75+80+92+87+85+92+93+88}{10} = \frac{827}{10} = 82.7$$

$$\Sigma(X_2 - \bar{X}_2)^2 = (4225+4900+5625+6400+8464+7569+7225+8464+8649+7744) - \frac{(827)^2}{10}$$

$$= 69265 - 68392.9 = 872.1$$

$$\therefore \sigma^2 = \frac{1833.6 + 872.1}{10+10-2} = \frac{2705.7}{18} = 150.3167 \quad \therefore \sigma = 12.26$$

$$t = \frac{97.2 - 82.7}{12.26} \sqrt{\frac{10 \times 10}{10+10}} = \frac{14.5}{12.26} \times 2.236 = \frac{32.422}{12.26} = 2.64$$

Degrees of freedom $= 10 + 10 - 2 = 18$

For 18 degrees of freedom, $t = 2.55$ is significant at ·02 level of significance. The observed value of t is greater than this and, therefore, highly significant. Therefore, it is unlikely that the two samples can be drawn at random from the same normal population.

PROBLEM 18.40. *A random sample of 16 men from district A has a mean height of 168.0 cms and a sum of squares from the sample mean of 132. A random sample of 25 men from district B had a mean height of 166.5 cms. and a sum of squares of 165. Can the samples be regarded as random samples from the same normal population ?*

Calculate the 95% fiducial limits of the difference between the population means.

SOLUTION.

First Sample	Second Sample
Size of the sample $= n_1 = 16$	Size of the sample $= n_2 = 25$
Sample mean $= \bar{X}_1 = 168.0$ cms.	Sample mean $= \bar{X}_2 = 166.5$ cms.
and $\Sigma(X_1 - \bar{X}_1)^2 = 132$	and $\Sigma(X_2 - \bar{X}_2) = 165$

$$\therefore \sigma^2 = \frac{132+165}{16+25-2} = \frac{297}{39} = \frac{99}{13} = 7.6158 \quad \therefore \sigma = \sqrt{7.6158} = 2{\cdot}76$$

Our hypothesis is that both the samples are from the same normal population.

$$t = \frac{168.0 - 166.5}{2.76} \sqrt{\frac{16 \times 25}{16+25}} = \frac{1.25}{2{\cdot}76} \times \frac{20}{6.4031} = \frac{25}{17.67} = 1.4 .$$

Degrees of freedom $= 41 - 2 = 39$

Usually tables do not provide the values of t for all degrees of freedom more than 30. We can find out the significance of the observed value of t by either of the following methods :

(*i*) We know that for large values of degrees of freedom, t tends to be distributed normally with the mean equal to zero and variance equal to one. The 5% level of significance for a normal variate $= 1.96$. The observed value of t is less than this and is, therefore, not significant.

(*ii*) From the tables we find :

For 30 degrees of freedom, ·05 level of value of $t = 2{\cdot}042$

For 40 degrees of freedom, ·05 level of value of $t = 2.021$

The observed value of t is 1·4 and, therefore, is not significant. No doubt is thrown on the hypothesis that they are random samples from the same normal population.

(*ii*) 95% fiducial limits of the difference between population means :

The calculation of ·05 level values of t is based upon the following formula :

If D is the P% value of normal deviate, then P% level of significance of t given by

$$t_p = D\left[\,1+\frac{1+D^2}{4\sigma}+\frac{3+16D^2+5D^4}{96\sigma^2}+\dots\dots\dots\right]$$ Where σ = degrees of freedom

$$\therefore\ t\,(.05) = 1.96\left[\,1+\frac{1+3.842}{4\times 39}+\frac{3+16\times 8.842+5\times 14.76}{96\times 39^2}\right]$$

$$= 1.96\left[\,1+\frac{4\cdot842}{156}+\frac{138.3}{146016}\right]\ 2\cdot02\ \text{(approx)}.$$

For 39 degrees of freedom 95% or ·05 level value of t=2.02

Now 95% fiducial limits of the difference between the population means are ven by

$$(\bar{X}_1-\bar{X}_2)\pm t\,(\cdot05)\sigma\sqrt{\frac{n_1+n_2}{n_1\ n_2}}$$

$\therefore$ The limits are : $(68\cdot0-66\cdot5)\pm 2\ 02\times 2.76\sqrt{\dfrac{41}{16\times 25}}$

$= 1.5\pm 5\cdot575\ \dfrac{6\cdot403}{20} = 1\cdot5\pm 1\cdot78$

i.e., from —0·28 to 3·28

Since zero lies inside this range, it further confirms that the difference etween samples means is not significant of a difference between the means of the opulations.

SIGNIFICANCE OF '*r*'

PROBLEM 18.41. *Apply t test to find whether correlation is significant, if* $= .6$ *and* $n=38$.

SOLUTION. Substituting the values of r and n in the formula

$$t=\frac{\sqrt{n-2}}{\sqrt{1-r^2}}\times r$$

Ve get.

$$t=\frac{\sqrt{38-2}}{\sqrt{1-(\cdot6)^2}}\times\cdot6=4\cdot5.$$

The number of degrees of freedom = 38—2, or 36. The value of t for 36 egrees of freedom at 5% level of significance is 2·025 approx. The calculated alue is considerably larger than this value, hence the correlation is significant.

PROBLEM 18·42. *A correlation co-efficient of +·5 is discovered in a sample* 9 *pairs. Is it significant ? Use t test.*

SOLUTION. Substituting the values of 'r' and 'n' in the formula

$$t=\frac{\sqrt{n-2}}{\sqrt{1-(r)^2}}\times r$$

Ve get,

$$t=\frac{\sqrt{19-2}}{\sqrt{1-(\cdot5)^2}}\times .5=27$$

The value of t for 17 degrees of freedom at 5% level is 2·11· The calculated value is higher than this and as such correlation is significant.

PROBLEM 18·43 *Use t test to find whether a correlation co-efficient of* $+\cdot 5$ *is significant if* $n=51$.

SOLUTION. Substituting the values of 'r' and 'n' in the formula

$$t=\frac{\sqrt{n-2}}{\sqrt{1-r^2}}\times r$$

We get,

$$t=\frac{\sqrt{51-2}}{\sqrt{1-(\cdot 5)^2}}\times\cdot 5=4\cdot 04$$

For 49 degrees of freedom at 5% level of significance the value of $t=2\cdot 01\cdot$ The calculated value being much higher that this the correlation is significant.

PROBLEM 18·44 *A random sample of 15 from a normal universe gives a correlation co-efficient of —0·5· Is this significant of the existence of correlation in the universe?*

SOLUTION. Substituting the values of 'r' and 'n' in the formula

$$t=\frac{\sqrt{n-2}}{\sqrt{1-r^2}}\times r$$

We get,

$$t=\frac{\sqrt{15-2}}{\sqrt{1-(-0\cdot 5)^2}}\times(-0\cdot 5)=-2\cdot 08$$

The number of freedom is 15—2, or 13. The value of t for 13 degrees of freedom at 5% level of significance is 2.16. The calculated value is considerably smaller than this figure. Hence, the correlation of the sample is not significant to warrant an existence of such correlation in the universe.

PROBLEM 18·45 *A random sample of 18 pairs of observations from a normal population gives a correlation coefficient of* $+0\cdot 52$. *Is it likely that the variables in the population are uncorrelated* ?

SOLUTION. We have $n=18$ and $r=+0\cdot 52$

$$\therefore t \quad \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

$$=\frac{0\cdot 52}{\sqrt{1-(0.52)^2}}\times\sqrt{18-2}$$

$$=\frac{0\cdot 52}{\sqrt{1-\cdot 2704}}\sqrt{16}=\frac{0\cdot 52\times 4}{\sqrt{0\cdot 7296}}=\frac{2\cdot 08}{0\cdot 85}=2\cdot 44$$

Degrees of freedom $=n-2=18-2=16$

For 16 degrees of freedom, $t=2.12$ is significant at ·05 level and $t=2.58$ is significant at ·02 level.

The observed value of t is significant at ·05 level but not significant at ·02 level. This gives reasons to suspect the hypothesis that the variables are uncorrelated. They may be correlated.

PROBLEM 18·46 *What is the least value of r in a random sample of 38 pairs that is significant (a) at the 0·05 level, (b) at the 0·01 level* ?

SOLUTION. Degrees of freedom$=n-2=38-2=36$

0·05 level value of the normal variate$=1.96$.

Now $t_p = D\left[1 + \frac{1+D^2}{4 \times \text{degrees of freedom}} + \ldots\ldots\right]$

When $P = .05$, $D = 1.96$; for 36 degrees of freedom, $t\,(0.05) = 1.96$ $[1 + \frac{4.842}{144} + \ldots\ldots] = 1.96 + 0.066 = 2.026$

0.01 level value of the normal variate $= 2.62$

When $P = .01$, $D = 2.62$; for 36 degrees of freedom, $t\,(0.01) = 2.62$ $[1 + \frac{7.864}{144} + \ldots\ldots] = 2.62 + 0.143 = 2.763$

If the value of r is significant at $p\%$, level, we have :

$$\frac{r}{\sqrt{1-r^2}}\sqrt{n-2} > t(p)$$

Substituting the values for n and t_p we find

(*a*) Value of r significant at .05 level is given by

$$\frac{r}{\sqrt{1-r^2}}\sqrt{36} \quad > 2.026$$

$$\text{or } 36\,r^2 \quad > 4.105 - 4.105r^2$$

$$\text{or } 40.105\,r^2 \quad > 4.105 \text{ or } r^2 > 0.1023\ r > 0.32$$

(*b*) Value of r significant at .01 level is given by

$$\frac{r}{\sqrt{1-r^2}}\sqrt{36} \ > 2.763 \text{ or } 36\,r^2 > 7.63 - 7.63\,r^2$$

$$\text{or } 43.63\,r^2 \quad > 7.63 \qquad \text{or } r^2 > 0.1748 \text{ or } r > 0.418$$

The least values of r significant at .05 or .01 levels are 0.32 and 0.418 respectively.

PROBLEM 18.47. *How many pairs of observations must be included in a sample in order that an observed correlation co-efficient of 0.42 shall have a calculated value of t greater than 2.72 ?*

SOLUTION. Let the number of pairs of observations be n

Then on substituting the values of r and t in the formula :

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

We have :

$$t > 2.72 \text{ or } t > \frac{0.42}{\sqrt{1-0.1764}}\sqrt{n-2}$$

$$\text{or } n-2 = \frac{0.8236}{0.1764} \times 7.3984 = 34.56 \quad \therefore n = 36.003 \text{ or } 36$$

There should be 36 pairs of observations if t is to be greater than 0.42.

PROBLEM 18.48. *Twelve pictures submitted in a competition were ranked by two judges with results as shown in the table below :*

Picture	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*J*	*K*	*L*
Ranks consigned by 1st Judge	5	9	6	7	1	3	4	12	2	11	10	8
Ranks consigned by 2nd Judge	5	8	9	11	3	1	2	10	4	12	7	6

Calculate P. Is there a lack of independence in these rankings? (Assume that on the hypothesis of independence of two sets of n rankings, $t = P\left(\frac{n-2}{1-p^2}\right)^{\frac{1}{2}}$ *follows the t-distribution with n—2 degree of freedom.)*

SOLUTION. We know that $P=1 \quad \frac{6\Sigma d^2}{n(n^2-1)}$, where n is the number of ranks, and d is the difference in ranks assigned to the same item.

In this problem $\Sigma d^2 = 0^2+1^2+3^2+4^2+2^2+2^2+2^2+2^2+2^2+1^2+3^2+2^2$

$$=1+9+16+4+4+4+4+4+1+9+4=60$$

and $n=12$

$$\therefore P=1-\frac{6\times 60}{12\,(144-1)}=1-\frac{30}{143}=1-0.2098=0.7902$$

Now, $t=P\,\frac{\sqrt{n-2}}{1-p^2}=0.7902\frac{\sqrt{12-2}}{\sqrt{1-0.624}}=\frac{0.7902}{\sqrt{0.3756}}\sqrt{10}$

$$=1.289\times 3.162=4.076$$

For 12—2=10 degrees of freedom, .05 level value of t is 2.29 and .01 level value of t is 3.17. The observed value of t is greater than this and therefore highly significant. The rankings indicate a lack of independence.

PROBLEM 18.49. *30 persons in the income group Rs. 20,000 to Rs. 25,000 were asked to supply before a specified date returns of their annual income for some reason connected with taxation. But when the date arrived, only 20 returns of total income, as noted below were received in the following order :*

20690,	*21720,*	*24010,*	*20090,*	*20940,*
21510,	*20340,*	*22420,*	*22180,*	*22600,*
21940,	*23080,*	*22840,*	*23510,*	*24260,*
23740,	*24720,*	*23310,*	*21300,*	*24530.*

Do the data confirm the belief that persons with bigger incomes delay submission of returns more than the others ? What is the measure of confidence you can associate with your answer. You may assume that the square of the test criterion t *is equal to* $\frac{(n-2)p^2}{1-p^2}$, *where P is the co-efficient of rank correlation and n the number of observations on which the co-efficient is based.*

SOLUTION.

Ranks (according to the Return in which returns were received).	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ranks (according to income).	3	7	17	1	4	6	2	10	9	11	8	13	12	15	18	16	20	14	5	19
d	2	5	14	3	1	0	5	2	0	1	3	1	1	1	3	0	3	4	14	1
d^2	4	25	196	9	1	0	25	4	0	1	9	1	1	1	9	0	9	16	196	1

$$P=1-\frac{6\Sigma d^2}{n(n^2-1)}=1-\frac{6\times 508}{20(400-1)}=1-\frac{3048}{7980}=\frac{4832}{7980}$$

$$=0.618$$

P=0.618 confirms the belief that persons with bigger incomes delay submission of returns.

Confidence will depend upon the level of significance of this value of the coefficient of correlation.

Given : $t^2=\frac{(n-2)\,p^2}{1-p^2}$ or $t=\frac{P}{\sqrt{1-P^2}}\sqrt{n-2}=\frac{0.618\times\sqrt{20-2}}{\sqrt{1-0.3819}}$

$$=\frac{0\,618\times 4.243}{0.786}=0.786\times 4.243=3.34$$

Degrees of freedom=20—2=18

For 18 degrees of freedom, $t=3.92$ is significant at .001 level. The observed value of t is less than this value and, therefore, not very highly significant, *i.e.*, it could have arisen due to chance factors only. [We can say with very great measure of confidence that there is no positive correlation between incomes and delay in submitting returns.]

PROBLEM 18.50. *A random sample of 15 from a normal universe gives a correlation of +.5. Use 'Z' test to find if it is significant of correlation in the universe.*

SOLUTION. Substituting the given value of 'r' in the formula

$$Z=\log_{10} \frac{(1+r)}{(1-r)} \times 1.1513$$

We get

$$Z=\log_{10} 3\times 1.1513.=.4771\times 1.1513.=.549$$

The standard error of Z

$$s.e._{Z}=\frac{1}{\sqrt{n-3}}=\frac{1}{\sqrt{15-3}}=.289 \text{ or } .29$$

If there is no correlation in the universe the value of ζ would be zero. The deviation of Z from ζ is .549 and it is less than twice the standard error of Z. The result is not decisive, though correlation can be said to be insignificant.

PROBLEM 18.51. *A correlation co-efficient of 0.5 is discovered in a sample of 19 pairs. Apply Z test to find out if this is significantly different.*

(*a*) *from 0.*

(*b*) *from 0.3.*

SOLUTION. Substituting the given value of 'r' in the formula

$$Z=\log._{10} \frac{1+r}{1-r} \times 1.1513$$

We get,

$$Z=\log._{10} \frac{1+.5}{1-.5} \times 1.1513=\log._{10} 3\times 1.1513=.4771\times 1.1513=.549.$$

The standard error of Z is

$$s.e._{Z}=\frac{1}{\sqrt{n-3}}=\frac{1}{\sqrt{19-3}}=.25$$

Now if we assume the correlation in the universe, *i.e.* p to be zero ζ is also equal to zero. The deviation of Z from ζ is more than twice the standard error of Z. Hence the hypothesis is wrong and we may conclude that the given correlation is significantly different from zero.

Note. For 'Z' test the co-efficient of correlation of a sample is converted into 'Z' by the formula $Z=\log_{10} \frac{1+r}{1-r} \times 1.1513$. Similarly the co-efficient of correlation of the universe or population is converted into ζ The formula used for the purpose is the same as above.

Now, if the supposition is that the correlation in the universe, *i.e.* $p=+0.3$. Then,

$$\zeta=\log_{10}\frac{1+p}{1-p} 1.1513.=\log_{10} \frac{1+.3}{1-.3} \times 1.1513$$

$$=\log_{10} 1.857\times 1.1513=.2688\times 1.1513=.309$$

The deviation of Z from ζ is, therefore,

0.309—0.549=—0.240.

This is only about .96 times of the standard error of Z, which is 0.25 and so could have arisen from the fluctuations of the sampling. Hence, it can be concluded that the given correlation co-efficient is not significantly different from +0.3.

PROBLEM 18.52. *Apply 'Z' test to determine the significance of difference between two correlations given by*

$r_1 = 0.6$, $r_2 = 0.5$
$n_1 = 9$, $n_2 = 12$

SOLUTION. If $r_1=0.6$.

$$Z_1=\log_{10}\frac{1+r}{1-r}\times 1.1513=\log_{10}\frac{1+0.6}{1-0.6}\times 1.1513$$

$$=\log_{10} 4\times 1.1513=.6021\times 1.1513=0.693.$$

Again, if $r_2=0.5$,

$$Z_2=\log_{10}\frac{1+r}{1-r}\times 1.1513=\log_{10}\frac{1+0.5}{1-0.5}\times 1.1513$$

$$=\log_{10} 3\times 1.1513=.4771\times 1.1513=0.549.$$

The difference of Z_1 and Z_2 is

$$Z_1-Z_2=0.693-0.549=0.144$$

The standard error of this difference of Z_1 and Z_2 is

$$s.e.\ z_1-z_2=\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}=\sqrt{\frac{1}{9-3}+\frac{1}{12-3}}=.53 \text{ approx}$$

The observed difference of the two Zs is only .27 times this standard error and so could be attributed to the sampling fluctuations. Hence the difference between the two correlations is not significant.

We can, however, calculate t at this stage.

$$t=\frac{Z_1-Z_2}{s.e.z_1-z_2}$$

Substituting the calculated values, we get,

$$t=\frac{0.693-0.549}{.53}=.27.$$

The number of degrees of freedom is N_1+N_2-6
$=9+12-6=15.$

The value of t for 15 degrees of freedom at 5% level of significance is 2.131. The calculated value is much less than this figure. Hence the difference is insignificant.

PROBLEM 18.53. *The correlation co-efficient between mathematics aptitude and language aptitude for a group of 20 boys is 0.42. For a group of 25 girls the correlation is 0.75. Is the difference significant?*

SOLUTION. If the correlation for the boys group, *i. e.* $r_1=0.42$.

$$Z_1=\log_{10}\frac{1+r}{1-r}\times 1.1513=\log_{10}\frac{1+0.42}{1-0.42}\times 1.1513$$

$$\log_{10} 2.448\times 1.1513=0.3888\times 1.1513=0.448$$

Similarly, if the correlation for the girls group, *i.e.* $r_2=0.75$,

$$Z_2=\log_{10}\frac{1+r}{1-r}\times 1.1513$$

$$=\log_{10}\frac{1+0.75}{1-0.75}\times 1.1513=\log_{10} 7\times 1.15$$

$$0.8451\times 1.1513=0.973.$$

The difference between the two Zs is

$Z_1 - Z_2 = 0.973 - 0.448 = 0.525$

The standard error of this difference is

$$s.e.\ z_1 - z_2 = \sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}} = \sqrt{\frac{1}{20-3}+\frac{1}{25-5}} = .33$$

The observed difference of the two Zs is about 1.6 times this standard error of the difference which may be therefore due to sampling fluctuations. Hence, we may conclude that the difference is not significant.

PROBLEM 18.54. *A random sample of 30 pairs of observations from a normal population shows a correlation co-efficient of 0.75. Is this consistent with the assumption that the correlation in the population is 0.55 ?*

SOLUTION. We have :

$$Z = \tfrac{1}{2}\log_{10}\frac{1+r}{1-r} \qquad \text{and } 3 = \tfrac{1}{2}\log_{10}\frac{1+P}{1-P}$$

$$= \tfrac{1}{2}\log_{10}\frac{1+0.75}{1-0.75} \qquad = \tfrac{1}{2}\log_{10}\frac{1+0.55}{1-0.55}$$

$$= \tfrac{1}{2}\times 2.3026\log_{10}\frac{1.75}{0.25} \qquad = \tfrac{1}{2}\times 2.3026\log_{10}\frac{1.55}{0.45}$$

$$= 1.1513\log_{10}7 \qquad = 1.1513\log_{10}3.444$$

$$= 1.1513\times 0.8451 \qquad = 1.1513\times 0.5371$$

$$= 0.9730 \qquad = 0.6194$$

Now, we know that Z is distributed normally with mean and standard deviation $\sqrt{\frac{1}{n-3}}$ *i. e.* $\frac{Z-\zeta}{\text{S. E.}}$ is a normal deviate

$$s.e._Z = \sqrt{\frac{1}{n-3}} = \sqrt{\frac{1}{30-3}} = \frac{1}{3\sqrt{3}} = \frac{\sqrt{3}}{9} = \frac{1.732}{9} = 0.1924$$

$$\therefore = \frac{Z-\zeta}{s.e.} = \frac{0{,}9730-0.6194}{0.1924}$$

$$\frac{0.3536}{0.1924} \doteq 1.833$$

.05 level value of normal deviate=1.96

Observed value of $\frac{Z-\zeta}{s.e.}$ is less than this and, therefore, it is not significant.

As far as the evidence goes the correlation in the population may be 0·55.

PROBLEM 18.55. *The value of r obtained from a random sample of 19 pairs of observations from a normal population is 0.8. Is this value consistent with the hypothesis that the correlation in the population is 0.6 ?*

SOLUTION. We know :

$$Z = \tfrac{1}{2}\log_{10}\frac{1+r}{1-r} \qquad = \tfrac{1}{2}\log_{10}\frac{1+P}{1-P}$$

$$= 1.1513\log_{10}\frac{1-0.8}{1-0.8} \qquad = 1.1513\log_{10}\frac{1-0.6}{1-0.6}$$

$$= 1.1513\log_{10}9 \qquad = 1.1513\log_{10}4$$

$$= 1.1513\times 0.9542 \qquad = 1.1513\times 0.6021$$

$$= 1.0986 \qquad = 0.6932$$

We know that Z is distributed normally with the mean and standard error $\sqrt{\frac{1}{n-3}}$ *i. e.* $\frac{Z-\zeta}{\text{S. E.}}$ is a normal deviate.

$$S.E.\ \frac{1}{\sqrt{19-3}} = \ = 0.25$$

$$\therefore \frac{Z-\zeta}{S.E.} = \frac{10986-0.6932}{0.25}$$

$$= \frac{0.4054}{0.25} = 1.6216$$

Observed value of $\frac{Z-\zeta}{S.E.}$ is less than this and, therefore, not significant.

The correlation in the normal population may be 0.6, as far as the evidence goes.

PROBLEM 18.56. *Two groups of students are given an intelligence test (x) and an arithmetic test.*

$n_1 = 45$ $\qquad r_1\ xy = 0.45$

$n_2 = 39$ $\qquad r_2\ xy = 0.38$

Is the difference between values of r significant.

SOLUTION.

First Sample	Second Sample
Size of the sample $= n_1 = 45$	Size of the sample $= n_2 = 39$
$r_1 xy = 0.45$	$r_2 xy = 0.38$
Now, $Z_1 = \frac{1}{2} \log_{10} \frac{1+r}{1-r}$	$Z_2 = \frac{1}{2} \log_{10} \frac{1+r}{1-r}$
$= 1.1513 \log_{10} \frac{1.45}{0.55}$	$1,1513 \log_{10} \frac{1.38}{0.62}$
$= 1.1513 \log_{10} 2.6364$	$= 1.1513 \log_{10} 2.2258$
$= 1.1513 \times 0.4207$	$= 1.1513 \times 0.3476$
$= 0.4844$	$= 0.4002$

If the samples are random samples from the same normal population $Z_1 - Z_2$ is distributed normally with mean zero and standard error

$$S.\ E. = \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}$$

$$S.\ E. = \sqrt{\frac{1}{45-3} + \frac{1}{39-3}}$$

$$= \sqrt{\frac{36+42}{42 \times 36}} = \frac{1}{6}\sqrt{\frac{78}{42}} = \frac{1}{6}\sqrt{\frac{13}{7}} = 0.227$$

$$\therefore \frac{Z_1 - Z_2}{s.\ e.} = \frac{0.484-0.400}{0.227} = \frac{0.084}{0.227} = 0.37$$

.05 level value of normal deviate 1.96.

The observed value is much less than this and, therefore. the difference between the values of *r* is not significant.

PROBLEM 18.57. *Two independent samples have 28 and 19 pairs of observations with correlation co-efficients 0.55 and 0.75 respectively. Are these values of r consistent with the hypothesis that the samples are drawn from the same population.*

Let $r_2 = 0.75$ within what limits must r_2 lie in order that $r_1 - r_2$ shall not be significant at the .05 level.

SOLUTION.

First Sample

$n_1 = 28$

$r_1 = 0.55$

$$Z_1 = \tfrac{1}{2}\ \log_{10} \frac{1+r_1}{1-r_1}$$

$$= 1.1513\ \log_{10} \frac{1.55}{0.45}$$

$$= 1.1513 \times 0.5371$$

$$= 0.6194$$

Second Sample

$n_2 = 19$

$r_2 = 0.75$

$$Z_2 = \tfrac{1}{2}\ \log_{10} \frac{1+r_2}{1-r_2}$$

$$= 1.1513\ \log_{10} \frac{1.75}{0.25}$$

$$= 1.1513 \times 0.8451$$

$$= 0.9730$$

If they are random samples from the same population $Z_1 - Z_2$ is distributed normally with the mean zero and standard error

$$\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

$$S.\ E. = \sqrt{\frac{1}{28-3} + \frac{1}{19-3}} = \sqrt{\frac{41}{25 \times 16}}$$

$$= \frac{6.403}{20} = 0.320$$

$$\therefore \frac{Z_1 - Z_2}{s.\ e.} = \frac{0.6194 - 0.9730}{0.32} = \frac{0.3546}{0.32} = 1.11$$

.05 level value of normal deviate $= 1.96$

The observed value is less than this and, therefore, not significant. There are no grounds to doubt the hypothesis.

(*ii*) Given the value of r_2, value of r_1 should be such that

$$\frac{Z_1 - Z_2}{s.\ e.} < 1.96$$

$$\text{or } \frac{Z_1 - 0.9730}{0.32} < 1.96 \text{ or } Z_1 - 0.9730 < 0.6272$$

This means : $Z_1 - 0.9730 < 0.6272$ or $Z_1 < 1.6002$

and $\quad 0.9730 - Z_1 < 0.6272$ or $Z_1 > 0.3458$

i.e. $\quad 0.3458 < Z_1 < 1.6002$

If the difference between r_1 and 0.75 is not to be significant at .05 level Z_1 must be within limits 0·3458 and 1·6002. Corresponding to these two limits of Z_1, we can find out the limits within which r_1 must lie.

$Z_1 > 0.3458$

or $1.1513 \log \dfrac{1+r_1}{1-r_1} > 0.3458$

or $\log \dfrac{1-r_1}{1-r_1} > 0.3004$

or $\dfrac{1-r_1}{1-r_1} > \text{Antilog } (0.3004)$

> 1.997

or $1 - r_1 > 1.997 - 1.997\, r_1$

or $2.997\, r_1 > 0.997$

or $\quad r_1 > 0.3326$

$Z_1 > 1.6002$

or $1.1513 \log \dfrac{1+r_1}{1-r_1} < 1.6002$

or $\log \dfrac{1-r_1}{1-r_1} < 1.3899$

or $\dfrac{1-r_1}{1-r_1} > \text{Antilog } (1.3899)$

> 24.54

or $1 - r_1 < 24.54 - 24.54\, r_1$

or $25.54\, r_1 < 23.54$

or $\quad r_1 < 0.9717$

Therefore, r_1 must be within the range 0.333 to 0.922 to *i. e.*, $0.922 < r_1 < 0.922$.

REVISIONAL PROBLEMS -18

PROBLEM 18.58. The Incomes of a random sample of engineers in industry I are : Rs. 630, 650, 680, 690. 720 and 710 per month. The incomes of a similar sample from industry II are Rs. 610, 620, 650, 660, 690, 690, 700, 710, 720 and 730 per month. Discuss the validity of the suggestion that industry I pays its engineers most better than industry II.

(Ans : t value = .1, hypothesis is correct).

PROBLEM 18.59. Eleven students of M. Com Final were given a test in Economic Analysis. They were given a month's special coaching and a second test of equal difficulty was held at the end of it. The results were as follows :

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks I Test	23	20	19	21	18	20	18	17	23	16	11
Marks II Test	24	19	22	18	20	22	20	20	23	20	17

Do the marks give an evidence that the students have benefited by extra coaching ? (When t for 10 degrees of freedom at 5% level of significance is 2.228.)

(Ans : $t=1.48$; students have not been benefited)

PROBLEM 18.60. Two kinds of manure were applied to 16 one acre plots, other conditions remaining the same, the yields (in qts.) are set out below : –

Manure I : 18, 20, 36, 50. 49, 36, 34, 49, 41 (9 plots)

Manure II : 29, 28, 26, 35, 30, 44, 46 (7 plots)

Examine the significance of the difference between the mean yields due to the applications of different kinds of manure. Use 't' test. (Ans : t (cal)=0.57)

PROBLEM 18.61. In two samples of 10 children, the increases in weight due to two different kinds (D_1 and D_2) in the same period in lbs were:

D_1	8	5	7	8	3	2	7	6	5	7
D_2	3	7	5	6	5	4	4	5	3	6

Find whether their variances and means differ significantly.

(Ans : $F=2.42$, $t=1.28$)

PROBLEM 18.62. A random sample of 30 pairs of observations from a normal population shows a correlation co-efficient of +.75. Is this consistent with the assumption that the correlation in the population is +0 55.

(Ans : $Z_1=0.97$; $Z_2=.62$; S. E.=0.19.)

PROBLEM 18.63. A random sample of 12 pairs of observations from a normal population gives a co-efficient of correlation of 0.45. Is this value significant of correlation in the population ? (Ans : $t=1.6$, not significant)

PROBLEM 18.64. In an examination in Psychology, 12 students in one class had a mean grade of 78 with a standard deviation of 6, while 15 students in another class had a mean grade of 74 with a standard deviation of 8. Is there a significant difference between the means of the two groups ?

(Ans : $t=1.65$, V=25, t *b* .05=2 06, No.)

PROBLEM 18.65. A random sample of 16 values from a normal population showed a mean of 41.5 and a sum of squares of deviations from the mean equal to 135. Can it be assumed that the mean of the population is 43.5. Use 5% level of significance. (Ans : $t=2.67$ for $v=15$ t .05=1.753)

PROBLEM 18.66. The means of the random samples of sizes 9 and 7 are 186.42 and 198.42 respectively. The sums of the squares of the deviations from the mean are 26.94 and 18.73 respectively. Can the samples be considered to have been drawn from the same normal population ?

(Ans: $t=2.21$; $v=14$; t 0.05 for v 14=2.145; No)

PROBLEM 18.67. To compare the price of a certain commodity in two towns, ten shops were selected at random in each town. The following figures give the prices found :

Town A	61	63	56	63	56	63	59	56	44	61
Town B	55	54	47	59	51	61	57	54	64	58

Test whether the average price can be said to be the same in two towns.

[Ans : $t=0.844$: $v=18$; for $v=18$, t .05=2.10
Average price is the same in two towns]

nalysis of Variance

19

The analysis of variance is a very useful device for analysing the results of ientific enquiries conducted in the field of biology and agriculture. Variance of a ries is split into 2 parts, *viz.*: (*i*) variance *between* varieties and (*ii*) variance *within* arieties. The variance between varieties is generally greater than the variance within e varieties. If there is a considerable difference between these variances then there ust be a specific cause for it. In order to test the significance of this difference "F" calculated. $F=\frac{V_2}{V_1}$ where V_2 and V_1 represent these variances between and within e varieties respectively. The calculated value of F is compared with its table value d if the calculated value is greater than the table value the difference is significant.

In this chapter questions about Fisher's Z-test have also been solved.

PROBLEM 19.1. *The figures given below are yields in bushels per acre of plots of gram. Three of these plots are of variety A and three of variety B.*

A	*30*	*32*	*22*
B	*20*	*18*	*16*

Set up a table of analysis of variance and calculate F to find out the significance f difference between the yields of two varieties.

SOLUTION. Calculation of the variance within the varieties and between the arieties.

Variety of gram	No. of Plots	Yield in each plot	Mean plot yield of varieties	Deviation of plot yield of each variety from its mean plot yield	Square of deviation	Deviation of mean plot yield of varieties from the mean plot yield of all varieties (23)	Square of deviation
A	1	30		+2	4		
	2	32	28	+4	16		
	3	22		−6	36	+5	25
B	1	20		+2	4		
	2	18	18	0	0		
	3	16		−2	4	−5	25
	6	138			64		50

From the above table

the sum of squares of deviations for within varieties=64 and the sum of squares of deviations for between varieties

$=50\times 3$ (*i.e.,*) the number of plots in a variety)

$=150.$

The number of degrees of freedom for within varieties

or $n_1=k\,(n-1)$; where k represents the number of varieties and n, the number of plots in each variety.

Substituting the values, we get,

$n_1=2\,(3-1)=4.$

The number of degrees of freedom for between varieties

or $n_2=(k-1)=2-1=1.$

With the help of these calculated values, we get the following table of analysis of variance.

Table of Analysis of Variance

Source of sum of Squares	Sum of Squares	Degrees of freedom	Mean square of variance
Within varieties	64	4	16
Between varieties	150	1	150
Total	214	5	

From the above table,

$$F=\frac{V_2}{V_1}=\frac{\text{Variance between varieties}}{\text{Variance within varieties}}=\frac{150}{16}=9.375.$$

The 5% value cf F for the calculated number of degrees of freedom is 7.71 which is considerably less than the calculated value. Hence, the variety differenc in yields is significant.

Alternate Method

Sample size		Square of deviations from grand average	Square of deviation of the mean of various samples from the grand average	Square of deviations of items from their sample average
A	30	49	25	4
	32	81	25	16
	22	1	25	36
B	20	9	25	4
	18	25	25	0
	16	49	25	4
			25	
Total		214	150	64

We can now have a table as the first method and obtain the value of F.

Short-cut method

Sample size		Square of items
A	30	900
	32	1024
	22	484
B	20	400
	18	324
	16	256
T=138		3388

Total sum of squares$=3388 \quad \frac{T^2}{N}$

$$=3388-\frac{138^2}{6}=3388-3174=214$$

Sum of squares between samples

$$=\frac{(\Sigma A)^2}{n_1}+\frac{(\Sigma B)^2}{n_2}-\frac{T^2}{N}$$

Where (ΣA) and (ΣB) stand for the total of items in the base samples A and nd n_1 and n_2 stand for the number of items in the two samples respectively. It l be equal to

$$=\frac{84^2}{3}+\frac{54^2}{3}-\frac{138^2}{6}=2352+972-3174=150$$

Sum of Squares within the samples

$=214--150=64$

Now we can set up a table of the analysis of variance as was done in the first ethod.

PROBLEM 19.2. *The following table gives the yields of four plots each of three rieties of wheat. Find out if the variety differences are significant.*

Plots	Plot yields				Totals
A	*26*	*27*	*28*	*31*	*112*
B	*18*	*19*	*21*	*22*	*80*
C	*16*	*18*	*19*	*19*	*72*
				Total	*264*

SOLUTION. Calculation of the variance within the varieties and between the rieties.

ariety of wheat	No. of Plots	Yield in each plot	Mean plot yield of varieties	Deviation of plot yield of each variety from its mean yield	Square of deviation	Deviation of mean plot yield of varieties from the mean plot yield of all varieties (22)	Square of deviation
A	1	26		−2	4		
	2	27		−1	1	+6	36
	3	28	28	0	0		
	4	31		+3	9		
B	1	18		−2	4		
	2	19		−1	1		
	3	21	20	+1	1	−2	4
	4	22		+2	4		
C	1	16		−2	4		
	2	18		0	0		
	3	19	18	+1	1	−4	16
	4	19		+1	1		
					30		56

From the above table,

the sum of squares of deviations for within varieties or $S_1=30$
the sum of squares of deviations for between varieties,

or $S_2=56\times4=224$.

The number of degrees of freedom for within varieties or is $n_1=k\,(n-1)$, where

k represents the number of varieties and n, the number of observations in each variety,

$n_1=3\,(4-1)=9$.

The number of degrees of freedom for between varieties, or $n_2=(k-1)=(3-1)=2$.

The variance for between varieties is

$$V_1=\frac{S_1}{n_1}=\frac{30}{9}=3.33$$

The variance for between varieties is

$$V_2=\frac{S_2}{n_2}=\frac{224}{2}=112$$

Substituting the values of V_1 and V_2 in the formula

$$F=\frac{V_2}{V_1}$$

We get,

$$F=\frac{112}{3.33}=33.6.$$

Using Snedecor's tables of F, we find that the 5% value of F for the given number of degrees of freedom is 4.26. Hence, we come to the conclusion that variety differences are quite significant.

PROBLEM 19.3. *Set up a table of analysis of variance for yields of three strains of wheat planted in five randomised blocks.*

Strains	*Blocks*				
A	*20*	*21*	*23*	*10*	*20*
B	*18*	*20*	*17*	*15*	*25*
C	*25*	*28*	*22*	*28*	*32*

SOLUTION. Calculation of the variance within the varieties and between the varieties.

Strains	Blocks	Yield of wheat in each block	Mean Block yield of strains	Deviation of Block yield of each strain from its mean block yield	Square of deviation	Deviation of mean block yield of strains from the mean block yield of all strains (22)	Square of deviation
	1	20		0	0		
	2	21		+1	1		
A	3	23	20	+3	9	−2	4
	4	16		−4	16		
	5	20		0	0		
	1	18		−1	1		
	2	20		+1	1		
B	3	17	19	−2	4	−3	9
	4	15		−4	16		
	5	25		+6	36		
	1	25		−2	4		
	2	28		+1	1		
C	3	22	27	−5	25	+5	25
	4	28		+1	1		
	5	32		+5	25		
					140		38

From the above table,

the sum of squares of deviations for within strains $=140$

and the sum of squares of deviations for between strains

$=38\times5=190$.

The number of degrees of freedom for within strains,

or $n_1=k\ (n-1)=3\ (5-1)=12$

Similarly, the number of degrees of freedom for between strains,

or $n_2=(k-1)=3-1=2$.

With the help of these calculated values, we get the following table of analysis of variance.

Table of Analysis of Variance

Source of sum of squares	Sum of squares	Degrees of freedom	Mean Square of variance	F
Within strains	140	12	11.67	8.14
Between strains	190	2	95	
Total	330	14		

From the table it is clear that the variance between strains is considerably high as compared to that for within varieties. The 5% value of F for the given number of degrees of freedom is 3.88 which is much less than the above calculated value. Hence the difference is significant.

PROBLEM 19.4. *Set up a table of analysis of variance for :—*

Plots	*Varieties*			
	a	*b*	*c*	*d*
1	*200*	*230*	*250*	*300*
2	*190*	*270*	*300*	*270*
3	*240*	*150*	*145*	*180*

SOLUTION. Calculation of the variance within the varieties and between the varieties.

Varieties	Plots	Yield of each plot	Mean plot yield of varieties	Deviation of plot yield of each variety from its mean plot yield	Square of deviation	Deviation of mean plot yield of varieties from the mean plot yield of all varieties (227)	Square of deviation
	1	200		−10	100		
a	2	190	210	−20	400	−17	289
	3	240		+30	900		
	1	230		+13	169		
b	2	270	217	+53	2809	−10	100
	3	150		−67	4489		
	1	250		+18	324		
c	2	300	232	+68	4624	+ 5	25
	3	145		−87	7569		
	1	300		+50	2500		
d	2	270	250	+20	400	+23	529
	3	180		−70	4900		
					29184		943

From the table,

the sum of the squares of deviations for within the varieties=29184.

and the sum of the squares of deviations for between the varieties $=943\times 3=2829$.

The number of degrees of freedom for within the varieties

$n_1=k\ (n-1)=4\ (3-1)=8$.

Similarly, the number of degrees of freedom for between the varieties is

$n_2=(k-1)=4-1=3$.

With the help of these calculated values, we set the table of the analysis of variance as follows :

Source of sum of squares	Sum of Squares	Degrees of freedom	Variances (sum of squares+d.f.)
Within varieties	29184	8	3648
Between varieties	2829	2	1414.5
Total	32013	10	

A perusal of the above table brings out clearly the fact that variance is more thin the varieties than between the varieties.

PROBLEM 19.5. *The following table gives the results of experiments on four rieties of a crop in 5 blocks of plots :—*

	Block				
	1	*2*	*3*	*4*	*5*
A	*32*	*34*	*33*	*35*	*37*
B	*34*	*33*	*36*	*37*	*35*
Variety C	*31*	*34*	*35*	*32*	*36*
D	*29*	*26*	*30*	*28*	*29*

Prepare the table of analysis of variance to test the significance of difference tween the yields of the four varieties.

SOLUTION. Calculation of the variance within the varieties and between the rieties.

Variety	No. of Blocks	Yields of each block	Mean block yields of varieties	Deviation of the block yield of each variety from its mean block yield	Square of deviation	Deviation of the mean block yield of each variety from the mean block yield of all varieties (32.8)	Square of deviation
	1	32		−2.2	4.84		
	2	34		−0.2	0.04		
A	3	33	34.2	−1.2	1.44	+1.4	1.96
	4	35		+0.8	0.64		
	5	37		+2.8	7.84		
	1	34		−1.0	1.00		
	2	33		−2.0	4.00		
B	3	36	35.0	+1.0	1.00	+2.2	4.84
	4	37		+2.0	4.00		
	5	35		0.0	0.00		
	1	31		−2.6	6.76		
	2	34		+0.4	0.16		
C	3	35	33.6	+1.4	1.96	+0.8	.64
	4	32		−1.6	2.56		
	5	36		+2.4	5.76		
	1	29		+0.6	0.36		
	2	26		−2.4	5.76	−4.4	19.36
D	3	30	28.4	+1.6	2.56		
	4	28		−0.4	0.16		
	5	29		+0.6	0.36		
					51.20		26.80

From the above table,

the sum of squares of deviations for within the varieties
=51.20

and the sum of squares of deviations for between the varieties
$=26.80 \times 5 = 134.00$

The number of degrees of freedom for within the varieties is

$n_1=k(n-1)=4\,(5-1)=16.$

and the number of degrees of freedom for between the varieties is

$n_2=(k-1)=4-1=3$

With the help of the above data, the table of the analysis of variance is set u as follows :

Source of sum of squares (1)	Sum of squares (2)	Degrees of freedom (3)	Variance (col. 2÷col. 3) (4)
Within varieties Between varieties	51.20 134.00	16 3	3.20 44.67
Total	185.20	19	

From the above table.

$$F=\frac{V_2}{V_1}=\frac{\text{Variance between varieties}}{\text{Variance within varieties}}=\frac{44.67}{3.20}=13.9$$

The 5% value of F for the above numbers of degrees of freedom is 3.24 whic is considerably less than the calculated value. Hence the difference between th yields due to varieties is significant.

PROBLEM 19.6. *The following table gives the strength of lead in Number o Pencils manufactured by "Company D."*

Test	*Pencil 1 Strength in kilograms*	*Pencil 2 strength in kilograms*	*Pencil 3 strength in kilograms*	*Pencil 4 strength in kilograms*	*Pencil 5 strength in kilograms*
1	*1.82*	*1.70*	*1.70*	*1.82*	*1.92*
2	*1.56*	*1.36*	*1.68*	*1.98*	*1.85*
3	*1.78*	*1.54*	*2.02*	*1.84*	*1.64*
4	*1.76*	*1.92*	*1.92*	*1.64*	*1.75*

Prepare the table of analysis of variance to test the significance of difference between the strength of pencils.

SOLUTION. Calculation of the variance within the varieties and between th varieties.

Variety of Pencil	No. of Blocks	Strength in each test in Kilograms	Mean strength of pencil in each variety	Deviations of the strength of pencil of each variety from its mean strength	Square of deviation	Deviation of the mean strength of each variety from the mean strength of all the varieties taken together (1.76)	Square of deviation
Pencil 1	1	1.82		+.09	.0081		
	2	1.56	1.73	−.17	.0289	−.03	.0009
	3	1.78		+.05	.0025		
	4	1.76		+.03	.0009		
Pencil 2	1	1.70		+.07	.0049		
	2	1.36	1.63	−.27	.0729	−.13	.0169
	3	1.54		−.09	.0081		
	4	1.92		+.29	.0841		

Pencil 3	1	1.70		—.13	.0169		
	2	1.68	1.83	—.15	.0225	+.07	.0049
	3	2.02		+.19	.0361		
	4	1.92		+.09	.0081		
Pencil 4	1	1.82		0	0		
	2	1.98	1.82	+.16	.0256	+.06	.0036
	3	1.84		+.02	.0004		
	4	1.64		—.18	.0324		
Pencil 5	1	1.92		+.13	.0169		
	2	1.85	1.79	+.06	.0036	+.03	.0009
	3	1.64		—.15	.0225		
	4	1.75		—.04	.0016		
					.3970		.0272

From the above table,

the sum of deviation for within varieties is 0.3970 :

and the sum of squares of deviations between the varieties is

.0272 × 3=0.0816.

The number of degrees of freedom for within varieties is

$n_1 = k\ (n—1) = 5\ (4—1) = 15.$

Similarly, the number of degrees of freedom between varieties is

$n_2 = (k—1) = 5—1 = 4.$

With the help of this data, the following table of analysis of variance is set up :

Table of Analysis of Variance

Source of sum of Squares	Sum of squares	Degrees of freedom	Variance
Within varieties	0.3970	15	0.0265
Between varieties	0.0816	4	0.0204
Total	0.4786	19	

The 5% value of F for the calculated numbers of degrees of freedom is 3.06. The calculated value will be considerably less than this figure. Hence, the difference between the strength of pencils of different varieties can be said insignificant.

PROBLEM 19.7. *The following table gives the length of Cuckoo's eggs deposited in the nests of three species of birds.*

Order of measurement	*Tree-pipit length in millimetres*	*Pied Wagtail Length in millimetres*	*Wren Length in millimetres*
1	*22.7*	*23.0*	*19.8*
2	*23.3*	*23.4*	*22.1*
3	*24.0*	*24.0*	*21.5*
4	*23.6*	*23.3*	*20.9*
5	*22.1*	*23.4*	*22.0*
6	*21.8*	*22.4*	*21.2*
7	*21.1*	*21.8*	*22.3*
8	*23.4*	*21.8*	*21.0*
9	*23.8*	*24.9*	*20.3*
10	*23.2*	*24.0*	*20.9*

Prepare the table of analysis of variance to test the significance of difference between the lengths of their species.

SOLUTION. Calculation of variance within the Nests and between the Nests.

Nests	Order of measurement	Length of the egg in each measurement in millimetres.	Mean length of egg in each nest (in millimetres).	Deviation of the length of the egg of each nest from the mean length	Square of the deviation	Deviation of the mean length of each nest from the mean length of all the nests taken together (22.4)	Square of deviation
Tree pipit	1	22.7	22.5	−0.2	0.04	+0.5	0.25
	2	23.3		+0.4	0.16		
	3	24.0		+1.1	1.21		
	4	23.6		+0.7	0.49		
	5	22.1		−0.8	0.64		
	6	21.8		−1.1	1.21		
	7	21.1		−1.8	3.24		
	8	23.4		+0.5	0.25		
	9	23.8		+0.9	0.81		
	10	23.2		+0.3	0.09		
Pied Wagtail	1	23.0	23.2	−0.2	0.04	+0.8	0.64
	2	23.4		+0.2	0.04		
	3	24.0		+0.8	0.64		
	4	23.3		+0.1	0.01		
	5	23.4		+0.2	0.64		
	6	22.4		−0.8	0.96		
	7	21.8		−1.4	1.96		
	8	21.8		−1.4	1.96		
	9	24.9		+1.7	2.89		
	10	24.0		+0.8	0.64		
Wren	1	19.8	21.2	−1.4	1.96	+1.2	1.44
	2	22.1		+9.0	0.81		
	3	21.5		+0.3	0.09		
	4	20.9		−0.3	0.09		
	5	22.0		+0.8	0.64		
	6	21.2		0	0		
	7	22.3		+1.1	1.21		
	8	21.0		−0.2	0.04		
	9	20.3		−0.9	0.81		
	10	20.9		−0.3	0.09		
					22.74		2.33

From the above table, the sum of the squares of deviation for within nests is 22.74 and the sum of the squares of deviations for between nests.

$=2.33\times 10=23.3$

The number of degrees of freedom for within nests is

$n_1=k\,(n-1)=3\,(10-1)=27.$

Similarly, the number of degrees of freedom for between nests is

$n_2=(k-1)\times 3-1=2$

With the help of the above data the following table of analysis of variance is set up.

Table of Analysis of Variance

Source of sum of Square	Sum of Squares	Degrees of freedom	Variance
ithin varieties	22.74	27	.842
etween varieties	23.30	2	11.65
Total	46.04	29	

Substituting the above values in the formula $F=\frac{V_2}{V_1}$ where 'represents the variance for between varieties and 'V_1' the variance for within the eties.

$$F=\frac{11.65}{.842}=13.83$$

The 5% value of F for the calculated numbers of degrees of freedom is 3.35. calculated value is many times greater than this value. Hence, the difference he length of the cuckoo's eggs due to difference in nests in which they were osited is significant.

PROBLEM 19.8. *Set up a table of analysis of variance and calculate F from results of the following four samples :—*

Sample 1	*Sample 2*	*Sample 3*	*Sample 4*
2	*3*	*6*	*5*
3	*4*	*8*	*5*
1	*3*	*7*	*5*
3	*5*	*4*	*3*
1	*0*	*10*	*2*

SOLUTION. Calculation of variance between and within the samples.

amples	Size	Mean of samples	Deviation of samples size from its mean	Square of deviation	Deviation of means of samples from combined mean (4)	Square of deviation
1	2		0	0		
	3		+1	1		
	1	2	−1	1		
	3		+1	1	−2	4
	1		−1	1		
2	3		0	0		
	4		+1	1		
	3	3	0	0	−1	1
	5		+2	4		
	0		−3	9		
3	6		−1	1		
	8		+1	1	+3	9
	7	7	0	0		
	4		−3	9		
	10		+3	9		
4	5		+1	1		
	5		+1	1		
	5	4	+1	1		
	3		−1	1	0	0
	2		−2	4		
				46		14

From the above we find that the sum of the squares of deviations within th samples=46.

The sum of the squares of deviations between the samples$=14\times5$ (number items in each sample)=70.

Number of degrees of freedom within the samples or

$n_1=k\,(n-1)$, where 'k' represents the number of samples and 'n' the number of items in each sample

$=4\,(5-1)=16.$

Number of degrees of freedom between the samples or

$n_2=(k-1)=(4-1)=3.$

Table of Analysis of Variance

Source of variation	Sum of squares	Degree of freedom	Mean square of variance
Within the Samples	46	16	$\frac{46}{16}=2.9$
Between the Samples	70	3	$\frac{70}{3}=23.3$

From the above table

$$F=\frac{\text{Variance between the samples}}{\text{Variance within the samples}}=\frac{V_2}{V_1}$$

$$=\frac{23.3}{2.9}=8.04$$

PROBLEM 19.9. *The following data relate to the results obtained by 4 investigators, investigating common problem. Each of the 4 investigators had taken a sample of 6 items. Do these results significantly vary from each other ?*

A	*B*	*C*	*D*
66	*42*	*54*	*78*
82	*66*	*90*	*54*
60	*30*	*60*	*60*
50	*60*	*81*	*42*
60	*36*	*60*	*71*
90	*48*	*51*	*49*

SOLUTION. Mean of the four samples are 68, 47, 66 and 59 respectively.

Deviations from respective means squared up

A	B	C	D
4	25	144	361
196	361	576	25
64	289	36	1
324	169	225	289
64	121	36	144
484	1	225	100
1136	966	1242	920

The sum of the squares of deviations within the samples thus

$=1136+966+1242+920=4264$

The combined mean $= 60$, and the squares of the deviations of sample means om it total to $64+169+36+1=270$

The total variance between the samples

$=270\times 6$ (number of items in each sample)$=1620$.

Degrees of freedom within the samples

or $n_1=k\ (n-1)=4\ (6-1)=20$

Degrees of freedom between the samples

or $n_2=(k-1)=3$.

Table of Analysis of Variance

Source of Variation	Sum of Squares	Degrees of Freedom	Mean of Square of Variance
Between Samples	1620	3	$\frac{1620}{3}=540$
Within the Samples	4264	20	$\frac{4264}{20}=213.2$
Total	5884	23	

$$F=\frac{V_2}{V_1}=\frac{540}{213.2}=2.5$$

For the degrees of freedom as in the problem at 5% level the value of $F=3.1$. e calculated value is less than this and as such the differences are not nificant.

PROBLEM 19.10. *Two independent samples of 8 and 7 items respectively had e following values of the variables (weight in grams).*

Sample 1 : 9 11 13 11 15 9 12 14
Sample 2 : 10 12 10 14 9 8 10

Do the two estimates of population variance differ significantly ?

SOLUTION.

Sample 1

x_1	x_1^2
9	81
11	121
13	169
11	121
15	225
9	81
12	144
14	196
94	1138

Sample 2

x_2	x_2^2
10	100
12	144
10	100
14	196
9	81
8	64
10	100
...	...
73	785

$$x_1 = \frac{94}{8} = 11.75 \qquad x_2 = \frac{78}{7} = 10.43$$

$$S_1^2 = \frac{1}{7}\left(1138 - \frac{94^2}{8}\right) = \frac{33.5}{7} \qquad S_2^2 = \frac{1}{6}\left(785 - \frac{73^2}{7}\right) = \frac{23.7}{6}$$

$$\text{Now } F = \frac{S_1^2}{S_2^2} = \frac{33.5}{7} \div \frac{23.7}{6} = \frac{33.5}{7} \times \frac{6}{23.7} = 1.22$$

Degrees of freedom : $n_1 = 8-1 = 7$, $n_2 = 7-1 = 6$

For $n_1 = 7$, $n_2 = 6$, .05 level value of F is 4.21 (approx). The observed valu being much less than it, the difference is very far from significant. The sampl may well be drawn from the same population.

PROBLEM 19.11. *For a random sample of 10 pigs fed on diet A the increas in weight in a certain period were :*

10, 6, 16, 17, 13, 12, 8, 14, 15, 9

For another random sample of 12 pigs, fed on diet B, the increases in the sam period were :

7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17

Show that the estimates of the population variance from the samples are n significantly different.

SOLUTION.

Sample 1

x_1	x_1^2
10	100
6	36
16	256
17	289
13	169
12	144
8	64
14	196
15	225
9	81
120	1560

Sample 2

x_2	x_2^2
7	49
13	169
22	484
15	225
12	144
14	196
18	324
8	64
21	441
23	529
10	100
17	289
180	3014

$$\bar{x}_1 = \frac{120}{10} = 12 \qquad \bar{x}_2 = \frac{180}{12} = 15$$

$$S_1^2 = \frac{1}{9}\left(1560 - \frac{120^2}{10}\right) \qquad S_2^2 = \frac{1}{11}\left(3014 - \frac{180^2}{12}\right)$$

$$= \frac{1}{9}\left(1560 - 1440\right) = \frac{120}{9} \qquad = \frac{1}{11}\left(3014 - 2700\right) = \frac{314}{11}$$

$$\therefore\ F = \frac{S_1^2}{S_2^2} = \frac{120/9}{314/11} = \frac{120}{314} \times \frac{11}{9} = 0.47$$

$$n_1 = 10-1 = 9, n_2 = 12-1 = 11$$

For $n_1=9$ and $n_2=8$, .05 level value of F is 3.23 and for $n_1=9$ and $n_2=12$, .05 level vaule of F is 3.07. The observed value of F is much less than these values and therefore the estimates of population variance from the samples are not significantly different.

PROBLEM 19.12. *A test was given to five students taken at random from the senior fifth form of each of the three grammar schools of a town. The individual scores are given in the following table. Is there any significant difference between the scores of the three schools ?*

Sample 1	*9*	*7*	*6*	*5*	*8*
Sample 2	*7*	*4*	*5*	*4*	*5*
Sample 3	*6*	*1*	*6*	*7*	*6*

SOLUTION. There is one criterion of variation, namely, schools. We have to determine if the variation due to this factor differs significantly from the variation 'within schools'.

	Sample 1		Sample 2		Sample 3	
	x_1	x_1^2	x_2	x_2^2	x_2	x_2^2
	9	81	7	49	6	36
	7	49	4	16	5	25
	6	36	5	25	6	36
	5	25	4	16	7	49
	8	64	5	25	6	36
$T=(\Sigma x_i)$	35		25		30	
$x_i=$	7		5		6	
$T_i^2=$	1225		625		900	
$\Sigma x^2=$		255		131		182

$$N=3\times5=15\ ;\ \Sigma T_i=T=90;\ x=\frac{90}{15}=6\ ;\ \Sigma T_i^2=2750\ ;\ \Sigma x^2=568$$

$$(i)\ \Sigma(x-\bar{x})^2=\Sigma x^2-\frac{T}{N}=568-\frac{90^2}{15}=568-540=28$$

$$(ii)\ \Sigma(x-\bar{x}_j)^2=\Sigma x^2-\frac{\Sigma T_i^2}{r_i}=568-\left(\frac{1225}{5}+\frac{625}{5}+\frac{900}{5}\right)=568-\frac{2750}{5}$$

$$=568-550=18$$

$$(iii)\ \Sigma n_i\,(\bar{x}_i-\bar{x})_2=\frac{T_i^2}{n_i}-\frac{T^2}{n_i}=\frac{2750}{5}-\frac{8100}{15}=550-540=10$$

Source of variation	*d.f.*	Sum of squares	Mean Square	F
Between Sample Means	$(3-1)=2$	10	5	3.33
Within Samples	$(15-3)=12$	18	1.5	
Total	14	28	—	

For $n_1=2$ and $n_2=12$, 0.5 level value of F is 3.88. The observed value of =3.33 is less than this value and therefore not significant. The test affords no idence against the hypothesis that the schools are similar.

PROBLEM 19.13. *Two independent random samples, each of 8 individuals provide the following data. Test the significance of the difference between the sample means.*

Height in cms. (x_1)	*163*	*164*	*165*	*165*	*166*	*166*	*167*	*168*
Height in cms. (x_2)	*169*	*166*	*167*	*167*	*166*	*168*	*168*	*169*

SOLUTION. If we reduce all the values of the variable by the same amount, it will not have any effect on the values of the sums of squares. In this example we are reducing all the values of the variable by 160. Then,

Sample 1

	x_1	x_1^2
	3	9
	4	16
	5	25
	5	25
	6	36
	6	36
	7	49
	8	64
$Ti=(\Sigma xi)=44$		
Ti^2	1936	
$\frac{Ti^2}{ni}$	242	
		260

Sample 2

x_2	x_2^2
9	81
6	36
7	49
7	49
6	36
8	64
8	64
9	81
60	
3600	
450	
	460

$N=2\times 8=16$

$\Sigma Ti=T=104$

$\bar{x}=\frac{104}{16}=6.5$

$\Sigma x^2=720$

$$\Sigma(x-\bar{x})^2=\Sigma x^2-\frac{T^2}{N}=720-\frac{10816}{16}=720-676=44$$

$$\Sigma(x-\bar{xi})^2=\Sigma x^2-\frac{\Sigma Ti^2}{ni}=720-(242+450)=720-692=28$$

$$\Sigma ni(\bar{xi}-\bar{x})^2=\frac{\Sigma Ti^2}{ni}-\frac{T}{N}=(242+450)-676=692-676=16$$

Source of variation	Sum of squares	*d.f.*	Mean square	F
Between sample means	16	(2—1)= 1	16	8
Within Samples	28	(16—2)=14	2	
Total	44	15	—	

For $n_1=1$, $n_2=14$, .05 level value of F is 4.60 and .01 level value of F is 8.86. The observed value of F is greater than .05 level value but smaller than .01 level value. Therefore the difference is not significant at .01 level of significance.

PROBLEM 19.14. *35 plots of approximately equal fertility were sown with 7 different varieties of wheat, 5 plots to each variety, the distribution of varieties among the plots being random. The following table gives the yields of grain in bushels per acre .the 7 columns corresponding to the different varieties. Do the date indicate a significant difference in the yields of the varieties ?*

Varieties

(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*	*(vii)*
13	*15*	*14*	*14*	*17*	*15*	*16*
11	*11*	*10*	*10*	*15*	*9*	*12*
10	*13*	*12*	*15*	*14*	*13*	*13*
10	*18*	*13*	*17*	*19*	*14*	*15*
12	*12*	*11*	*10*	*12*	*10*	*11*

SOLUTION. In this example there is one criterion of variation, namely, varieties of wheat. To simplify the arithmetic we reduce all yields by 12. This table is given below.

	1	3	2	2	5	3	4
	−1	−1	−2	−2	3	−3	0
	−2	1	0	3	2	1	1
	4	6	1	5	7	2	3
	0	0	−1	−2	0	−2	−1
$T_i=$	2	9	0	6	17	1	7
$\Sigma x_i^2=$	22	47	10	46	87	27	27
$T_i^2=$	4	81	0	36	2.89	1	49
$\frac{T_i^2}{n_i}=$	0.8	16.2	0	7.2	57.8	0.2	9.8

$N=7\times5=35$

$\Sigma T_i=T=42$

$\bar{x}=\frac{42}{33}=1.2$

$\Sigma x^2=266$

$\sum\frac{T_i^2}{n_i}=92.0$

$$\Sigma(x-\bar{x})^2=\Sigma x^2-\frac{T^2}{N}=266-\frac{42^2}{35}=266-50.4=215.6$$

$$\Sigma(x-\bar{x})^2=\Sigma x^2-\frac{\Sigma T_i^2}{n_i}=266-92 \qquad =174$$

$$\Sigma n_i(\bar{x}_i-\bar{x})^2=\frac{\Sigma T_i^2}{N}-\frac{T^2}{N}=92-50.4 \quad =41.6$$

Source of Variation	Sum of Squares	D.F.	Mean square	F
Between varieties	41.6	7−1=6	6.933	1.1
Within varieties	174.0	35−7=28	6.214	
Total	215.6	34	—	

For $n_1=6$ and $n_2=25$; .05 level value of F is 2.49

For $n_1=6$ and $n_2=30$; .05 level value of F is 2.42

The observed value of F is less than the theoretical value of F and therefore not significant. The experiment does not show significant variation in the yields of varieties.

PROBLEM 19.15. *Use the data of above example tc consider the hypothesis that the four types do not differ significantly in gaining weight and that they were fed on four different rations and they also do not differ.*

SOLUTION. In this example we have two criteria —namely, type of pig (cols) and type of ration (rows). Therefore, we shall have to test row means as well as column means.

Type of pig

		A	B	C	D	T	$T_i{}^2$
	I	6.1	14.6	11.5	13.4	45.6	2079.36
	II	13.8	15.7	16.0	20.2	65.7	4316.49
Ration	III	8.7	11.8	9.0	12.9	42.4	1797.76
	IV	12.0	16.5	13.3	12.5	54.3	2948.49
T_i		40.6	58.6	49.8	59.0	208.0	11142.10
T_i^2		1648.36	3433.96	2480.04	3481.00	$N=4\times4=16$ $T=208\ \ \frac{T^2}{N}=2704$	

$$\Sigma(x-\bar{x})\Sigma x^2-\frac{T^2}{N}=2874.88-2704=170.88$$

$$\Sigma n_i(x_i-\bar{x})^2=\frac{\Sigma T_i{}^2}{n_i}-\frac{T^2}{N}=2760.84-2704=56.84$$

$$\Sigma n_i(\bar{x}_i-\bar{x})^2=\Sigma\frac{T_i^2}{n_i}-\frac{T^2}{N}=\frac{11142.1}{4}-2704=2785.525-2704=81.525$$

Source of Variation	*D.F.*	Sum of squares	Mean square	F
Columns	3	56.840	18.947	5.24
Rows	3	81.525	27.175	7.52
'Error' or Residue variance	9	32.515	3.613	—
Total	15	170.880	—	—

For $n_1=3$ and $n_2=9$, .05 level value of F is 3.86 and .01 level value is 6.99. Therefore the differences between column means are significant and the differences between row means are highly significant.

PROBLEM 19.16. *Two random samples drawn from normal populations are :*

Sample 1 : 20, 16, 26, 27, 23, 22, 18, 24, 25, 19

Sample 2 : 27, 33, 42, 35, 32, 34, 38, 28, 41, 43, 30, 37

Obtain estimates of the variances of the population and test whether the two populations have the same variance.

SOLUTION. Number of items in the first sample=10

$$\text{Mean of the first sample}=\frac{20+16+26+27+23+22+18+24+25+19}{10}$$

$$=\frac{220}{10}=22$$

Number of items in the second sample=12

Mean of the second sample

$$=\frac{27+33+42+35+32+34+38+28+41+43+30+37}{12}=\frac{420}{12}=35$$

First Sample

Deviations from mean

(x_2-22)	$(x_2-22)^2$
-2	4
-6	36
$+4$	16
$+5$	25
$+1$	1
0	0
-4	16
$+2$	4
$+3$	9
-3	9
0	120

Second Sample

Deviations from mean

(x_1-35)	$(x_1-35)^2$
-3	64
-2	4
$+7$	49
0	0
3	9
-1	1
$+3$	9
-7	49
$+6$	36
$+8$	64
-5	25
$+2$	4
0	314

$$S_2^2=\frac{120}{1-10}=\frac{120}{9}=13.33$$

$$n_2=10-1=9$$

$$S_1^2=\frac{314}{12-1}=\frac{314}{1}=28.55$$

$$n_1=12-1=11$$

$$\therefore F=\frac{S_1^2}{S_2^2}=\frac{28.55}{13.33}=2.24$$

For $n_1=8$ and $n_2=9$, .05 level value of F is 3.23. For $n_1=12$ and $n_2=9$, .05 level value of F is 3.07. The observed value of F $(=2.14)$ is less than this and therefore not at all significant. The sample may well have been taken from a population with the same variance.

PROBLEM 19.17. *The price of a certain commodity was ascertained in each of four towns and at each of four dates, one in each quarter of the year. Prices, in rupees are shown in the table. Are the variations between the different localities and between the different seasons significant ?*

Towns

Quarter	*A*	*B*	*C*	*D*
I	*6*	*5*	*6*	*5*
II	*5*	*4*	*6.5*	*5*
III	*4.5*	*3.5*	*4.5*	*5*
VI	*6.5*	*4.5*	*6*	*7*

SOLUTION. In this example there are two criteria of variations in towns and seasons. We reduce all the prices by 5. Then the table reads as follows :

Towns

Quarters	A	B	C	D	T_i	T_{i2}
I	1	0	1	0	2.0	4.00
II	0	-1	1.5	0	0.5	0.25
III	-0.5	-1.5	-0.5	0	-2.5	6.25
IV	1.5	-0.5	1	2	4.0	6.00
$T_j=$	2.0	-3.0	3.0	2	4	26.50
$T_j^2=$	4.0	9.0	9.0	4.0	26	

$T^2=16$; $N=16$; $x^2=1+0.25+2.25+1+2.25+.25+1+2.25+.25+1+4$

$=15.50$

$$\Sigma(x-\bar{x})^2=\Sigma x^2-\frac{T^2}{N}=15.50-\frac{16}{16}=14.50$$

$$\Sigma n_i(x_i-\bar{x})^2=\frac{\Sigma T_i^2}{n_i}-\frac{T^2}{N}=\frac{26}{4}-\frac{16}{16}=6.5-1=5.5$$

$$\Sigma n_i(\bar{x}_i-\bar{x})^2=\frac{\Sigma T_i^2}{n_i}-\frac{T^2}{N}=\frac{26.50}{4}-\frac{16}{16}=6.625-1=5.625$$

Source of Variation	Sum of squares	D.F.	Mean squares	F
Columns	5.5	4—1=3	1.833	4.89
Rows	5.625	4—1=3	1.875	5.00
Error or Residual	3.375	9	0.375	
Total	14.500	15		

For $n_1=3$ and $n_2=9$, .05 level value of $F=3.86$ and .01 level value of $F=6.99$. Both the observed values of F are greater than 3.86 but less than 6.99. Hence the variations are significant at .05 but not significant at .01 level.

Fisher's *Z*-test

PROBLEM 19.18. *In a class of 20 children, 10 chosen at random were given a ration of orange juice each day for a certain period and other ten a ratin milk. Their gains in weight during the period were : in pounds*

First group :	4	$2\frac{1}{2}$	$3\frac{1}{2}$	4	$1\frac{1}{2}$	1	$3\frac{1}{2}$	3	$2\frac{1}{2}$	3.5
Second group :	$1\frac{1}{2}$	$3\frac{1}{2}$	$2\frac{1}{2}$	3	$2\frac{1}{2}$	2	2	$2\frac{1}{2}$	$1\frac{1}{2}$	3

Apply Z-test to find out whether the sample variances indicate any significant difference in the present population ?

Extracts from statistical Tables

5% and 1% points of Z

n_1	8	12
n_2		
8	.6175	.5945
	.8983	.8673
9	.5662	.5613
	.8494	.8157
	.5611	.5346
10	.8104	.7744

SOLUTION. Mean of the first group

$$=\frac{4+2.5+3.5+4+1.5+1+3.5+3+2.5+3.5}{10}=\frac{29.0}{10}=2.9$$

Mean of the second group

$$=\frac{1.5+3.5+2.5+3+2.5+2+2+2.5+1.5+3}{10}=\frac{24.0}{10}=1.4$$

First Group		Secoad Group	
$(x_1-2.9)$	$(x_1-2.9)^2$	$(x_2-2.4)$	$(x_2-2.4)^2$
1.1	1.21	−0.9	0.81
−0.4	0.16	1.1	1.21
0.6	0.36	0.1	0.01
1.1	1.21	0.6	0.36
−1.4	1.96	0.1	0.01
−1.9	3.61	−0.4	0.16
0.6	0.36	−0.4	3.16
0.1	0.01	0.1	0.01
−0.4	0.16	−0.9	0.81
0.6	0.36	0.6	0.36
0.0	9.40	0.0	3.90

$(x-2.9)^2=9.40,\qquad n_1=10-1=9$

$(x-2.4)^2=3.90,\qquad n_2=10-1=9$

$S_1^2=\frac{9.40}{9}\ ;\qquad S_2^2=\frac{3.90}{9}$

$Z=\text{loge}\ \frac{S_1^2}{S}=\frac{1}{2}\log_e\ \frac{S_1^2}{S_2^2}=\frac{1}{2}\text{loge}\frac{9.40}{4}+\frac{3.90}{9}$

$=\frac{1}{2}(2.3026)\ \log_{10} 2.41=1.1513\times0.3820=0.43979=0.4398$

For $n_1=8$ and $n_2=9$, .05 leval value of Z is 0.5862

For $n_1=12$ and $n_2=9$, .05 level value of Z is 0.5613

The observed value of Z is than these values and therefore not significant.

PROBLEM 19.19. *The heights of six randomly chosen sailors are in cms 163,165,168.169,171 and 172. Those of ten randomly chosen soldiers are 161, 162,165, 166 169,169,170,171,172,173, Use the Z-distribution to discuss whether the samples can have come from population which are identical so far as height distribution is concerned.*

SOLUTION.

Second Sample Mean $=\frac{163+165+168+169+171+172}{6}$ and First Sample Mean $\frac{161+162+165+166+169+169+170+171+172+173}{10}$

$=\frac{1008}{6}=168;\ n_2=6-1=5$ and $=\frac{1678}{10}=167.8;\ n_1=10-1=9$

Deviations from mean		Deviations from mean	
(x_2-168)	$(x_2-168)^2$	$(x_1-167.8)$	$(x_1-167.8)$
−5	25	−6.8	46.24
−3	9	−5.8	33.64
0	0	−2.8	7.84
+1	1	−1.8	3.24
+3	9	1.2	1.44
+4	16	1.2	1.44
0	60	2.2	4.84
		3.2	10.24

4.2	17.64
5.2	27.04
0.0	153.60

$$\therefore S_2^2 = \frac{60}{6-1} = 12 \qquad S_1^2 = \frac{153.6}{10-1} = 17.07$$

$$Z = \log_e \frac{S_1}{S_2} = \tfrac{1}{2} \log_e \frac{S_1^2}{S_2^2} = \tfrac{1}{2} \log_e \frac{17.07}{12}$$

$$= \tfrac{1}{2} \log_e 1.4225 = \tfrac{1}{2} \times 0.3524 = 0.1762$$

For $n_1 = 8$ and $n_2 = 5$, .05 level value of $Z = 0.7862$

$n_1 = 12$ and $n_2 = 5$, .05 level value of $Z = 0.7714$

The observed value of Z is smaller than this and therefore not significant. The samples may well have been taken from a population with the same variance.

REVISIONAL PROBLEMS—19

PROBLEM 19.20. Below are given yields of three strains of wheat planted in blocks of 3 plots each under a completely randomized design. All the 15 plots are of equal size. Perform the analysis of variance to test whether the strains are significantly different with regard to yield. Ignore variation between blocks.

		BLOCKS				
		I	II	III	IV	V
Strain	A	20	21	23	16	20
	B	18	20	17	15	25
	C	25	28	22	28	32

(*Ans.* "*F*" *ratio=8.14.m*)

PROBLEM 19.21. The data below represent the throughputs obtained from units of an acid paint before failure through correction. The units are classified according to the foundry manufacturing them.

Foundry	Throughput obtained							
A	84	60	40	47	34			
B	67	92	95	40	98	60	59	109	86
C	46	93	100					

Perform an analysis of variance of these data.

(*Ans.* "*F*" *2.2, difference not significant*)

PROBLEM 19.22. In one sample of 8 observations, the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations it was 102.6. Test whether this difference is significant at 5 percent level, given that 5 percent point of F for $V_1 = 7$ and $V_2 = 9$ degrees of freedom is 3.29.

(*Ans.* "*F*" *= 1.06, difference not significant*)

PROBLEM 19.23. In an experiment conducted on a farm in a certain village of Rajasthan, the following information was collected regarding the yield in quintals per acre of 6 plots of wheat. Three out of these six plots produced Sharbati wheat and the rest three produced Australian wheat. Set up an analysis table and find out if the variety differences are significant. Use values of '*F*' at 5% level of significance.

Variety	Plot yield in Quintals		
Sharbati	10	15	11
Australian	13	12	17

(The value of "F" at 5% level of significance for $V_1=4$ and $V_2=1$ is 7.71)

(Ans. "F"=1.17, Difference not significant)

PROBLEM 19.24. The following figures relate to production in kg. of three varieties A, B and C of wheat sown in 12 plots—

A	14,	16,	18		
B	14,	13,	15,	22	
C	18,	16,	19,	19,	20

Is there any significant difference in the production of the three varieties.

(Ans. Between samples=17.2
within samples=67.20)

PROBLEM 19.25. The X Co. wishes to test whether its 3 salesmen A, B and C tend to make sales of the same size or whether they differ in their selling ability as measured by the average size of their sales. During the last week there have been 14 sale calls—A made 5 calls, B made 4 calls and C made 5 calls.

Following are the weekly sales record of the three salesmen :

A Rs.	B Rs.	C Rs.
300	600	700
400	300	300
300	300	400
500	400	600
0	—	500

Perform the analysis and draw your conclusions.

(Ans. within=16.5)

PROBLEM 19.26. To assess the significance of possible variation in performance in a certain test as between grammer schools of a city, a common test was given to a number of 5 students taken at random from the senior fifth form of each of the four schools concerned. The results are given below. Find out analysis of variance of data :

SCHOOL

A	B	C	D
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15

$$\left(Ans.\ \frac{V_2}{V_1}=\frac{16.7}{13}=1.28\right)$$

PROBLEM 19.27. A tea company appoints four salesmen A, B, C and D and observes their sales in 3 seasons—summer, winter and monsoon. The figures in (lakhs) are given below :

Seasons	Salesmen				Totals
	A	B	C	D	
Summer	36	36	21	35	128
Winter	28	29	31	32	120
Monsoon	26	28	29	29	112
Totals	90	93	81	96	360

Carry out analysis of variance.

(*Ans.* *"F"=1.62 Between Salesmen*
"F"=1.42 Between Seasons)

PROBLEM 19.28. The following table gives the yield of four plots each of three varieties of wheat. Find out if the variety differences are significant.

Plots	Plot yields				Total
A	52	54	56	62	224
B	36	38	42	44	160
C	32	36	38	38	144
					Total 528

$$\left(\textit{Ans. } \text{"}F\text{"} = \frac{V_2}{V_1} = \frac{448}{13.33} = 33.6 \right)$$

PROBLEM 19.29. Set up a table of analysis of variance and calculate *F* from the results of the following samples :—

Sample 1	Sample 2	Sample 3	Sample 4
4	6	12	10
6	8	16	10
2	6	14	10
6	10	8	6
2	5	20	4

MATHEMATICAL TABLES

Logarithms

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7

Logarithms

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	6	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	3	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	2	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4

ANTILOGARITHMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
·00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
·01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
·02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
·03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
·04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
·05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
·06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
·07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
·08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
·09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
·10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
·11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
·12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
·13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
·14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
·15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
·16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
·17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
·18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
·19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
·20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
·21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
·22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
·23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
·24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
·25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
·26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
·27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
·28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
·29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
·30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
·31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
·32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
·33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
·34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
·35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
·36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
·37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
·38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
·39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
·40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
·41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
·42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
·43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
·44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
·45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
·46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
·47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
·48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
·49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

ANTILOGARITHMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
·50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
·53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
·57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
·59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
·61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
·66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
·67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
·69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
·71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
·73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
·77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
·79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
·80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
·84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
·85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
·86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
·87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
·88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
·89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
·90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
·91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
·92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
·93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
·94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
·95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
·96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
·97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
·98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
·99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

POWERS, ROOTS & RECIPROCALS

n	n^2	n^3	$\sqrt{n}$	$\sqrt[3]{n}$	$\frac{1}{n}$	$\sqrt{10n}$	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{10n}}$
1	1	1	1·0000	1·0000	1·00000	3·1623	1·00000	0·31623
2	4	8	1·4142	1·2599	0·50000	4·4721	0·70711	0·22361
3	9	27	1·7321	1·4422	0·33333	5·4772	0·57735	0·18257
4	16	64	2·0000	1·5874	0·25000	6·3245	0·50000	0·15811
5	25	125	2·2361	1·7100	0·20000	7·0711	0·44721	0·14142
6	36	216	2·4495	1·8171	0·16667	7·7460	0·40825	0·12910
7	49	343	2·6458	1·9129	0·14286	8·3666	0·37796	0·11952
8	64	512	2·8284	2·0000	0·12500	8·9443	0·35355	0·11180
9	81	720	3·0000	2·0801	0·11111	9·4868	0·33333	0·10541
10	100	1000	3·1623	2·1544	0·10000	10·0000	0·31623	0·10000
11	121	1331	3·3166	2·2240	0·09091	10·4881	0·30151	0·09535
12	144	1728	3·4641	2·2894	0·08333	10·9545	0·28868	0·09129
13	169	2197	3·6056	2·3513	0·07692	11·4018	0·27735	0·08771
14	196	2744	3·7417	2·4101	0·07143	11·8322	0·26726	0·08452
15	225	3375	3·8730	2·4662	0·06667	12·2474	0·25820	0·08165
16	256	4096	4·0000	2·5198	0·06250	12·6491	0·25000	0·07906
17	289	4913	4·1231	2·5713	0·05882	13·0384	0·24253	0·07670
18	324	5832	4· 126	2·6207	0·05556	13·4164	0·23570	0·07454
19	361	6859	4·3589	2·6684	0·05263	13·7840	0·22942	0·07255
20	400	8000	4·4721	2·7144	0·05000	14·1421	0·22361	0·07071
21	441	9261	4·5826	2·7589	0·04762	14·4914	0·21822	0·06901
22	484	10648	4·6904	2·8020	0·04545	14·8324	0·21320	0·06742
23	529	12167	4·7958	2·8439	0·04348	15·1658	0·20851	0·06594
24	576	13824	4·8990	2·8845	0·04167	15·4919	0·20412	0·06455
25	625	15625	5·0000	2·9240	0·04000	15·8114	0·20000	0·06325
26	676	17576	5·0990	2·9625	0·03846	16·1245	0·19612	0·06202
27	729	19683	5·1962	3·0000	0·03704	16·4317	0·19245	0·06086
28	784	21952	5·2915	3·0366	0·03571	16·7332	0·18898	0·05976
29	841	24389	5·3852	3·0723	0·03448	17·0294	0·18570	0·05872
30	900	27000	5·4772	3·1072	0·03333	17·3205	0·18257	0·05774
31	961	29791	5·5678	3·1414	0·03226	17·6068	0·17961	0·05680
32	1024	32768	5·6569	3·1748	0·03125	17·8885	0·17678	0·05590
33	1089	35937	5·7446	3·2075	0·03030	18·1659	0·17408	0·05505
34	1156	39304	5·8310	3·2396	0·02941	18·4391	0·17150	0·05423
35	1225	42875	5·9161	3·2711	0·02857	18·7083	0·16903	0·05345
36	1296	46656	6·0000	3·3019	0·02778	18·9737	0·16667	0·05270
37	1369	50653	6·0828	3·3322	0·02703	19·2354	0·16440	0·05199
38	1444	54872	6·1644	3·3620	0·02632	19·4936	0·16222	0·05130
39	1521	59319	6·2450	3·3912	0·02564	19·7484	0·16013	0·05064
40	1600	64000	6·3245	3·4200	0·02500	20·0000	0·15811	0·05000
41	1681	68921	6·4031	3·4482	0·02439	20·2485	0·15617	0·04939
42	1764	74088	6·4807	3·4760	0·02381	20·4939	0·15430	0·04880
43	1849	79507	6·5574	3·5034	0·02326	20·7364	0·15250	0·04822
44	1936	85184	6·6332	3·5303	0·02273	20·9762	0·15076	0·04767
45	2025	91125	6·7082	3·5569	0·02222	21·2132	0·14907	0·04714
46	2116	97336	6·7823	3·5830	0·02174	21·4476	0·14744	0·04663
47	2209	103823	6·8557	3·6088	0·02128	21·6795	0·14587	0·04613
48	2304	110592	6·9282	3·6342	0·02083	21·9089	0·14434	0·04564
49	2401	117649	7·0000	3·6593	0·02041	22·1359	0·14286	0·04518
50	2500	125000	7·0711	3·6840	0·02000	22·3607	0·14142	0·04472

POWERS, ROOTS & RECIPROCALS

n	n^2	n^3	$\sqrt{n}$	$\sqrt[3]{n}$	$\frac{1}{n}$	$\sqrt{10n}$	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{10n}}$
51	2601	132651	7·1414	3·7084	0·01961	22·5832	0·14003	0·04428
52	2704	140608	7·2111	3·7325	0·01923	22·8035	0·13868	0·04385
53	2809	148877	7·2802	3·7563	0·01887	23·0217	0·13736	0·04344
54	2916	157464	7·3485	3·7798	0·01852	23·2379	0·13608	0·04303
55	3025	166375	7·4162	3·8030	0·01818	23·4521	0·13484	0·04264
56	3136	175616	7·4833	3·8259	0·01786	23·6643	0·13363	0·04226
57	3249	185193	7·5498	3·8485	0·01754	23·8747	0·13245	0·04189
58	3364	195112	7·6158	3·8709	0·01724	24·0832	0·13131	0·04152
59	3481	205379	7·6811	3·8930	0·01695	24·2899	0·13019	0·04117
60	3600	216000	7·7460	3·9149	0·01667	24·4949	0·12910	0·04082
61	3721	226981	7·8102	3·9365	0·01639	24·6982	0·12804	0·04049
62	3844	238328	7·8740	3·9579	0·01613	24·8998	0·12700	0·04016
63	3969	250047	7·9373	3·9791	0·01587	25·0998	0·12599	0·03984
64	4096	262144	8·0000	4·0000	0·01563	25·2982	0·12500	0·03953
65	4225	274625	8·0623	4·0207	0·01538	25·4951	0·12403	0·03922
66	4356	287496	8·1240	4·0412	0·01515	25·6905	0·12309	0·03892
67	4489	300763	8·1854	4·0615	0·01493	25·8844	0·12217	0·03863
68	4624	314432	8·2462	4·0817	0·01471	26·0768	0·12127	0·03835
69	4761	328509	8·3066	4·1016	0·01449	26·2679	0·12039	0·03807
70	4900	343000	8·3666	4·1213	0·01429	26·4575	0·11952	0·03780
71	5041	357911	8·4261	4·1408	0·01408	26·6458	0·11868	0·03753
72	5184	373248	8·4853	4·1602	0·01389	26·8328	0·11785	0·03727
73	5329	389017	8·5440	4·1793	0·01370	27·0185	0·11704	0·03701
74	5476	405224	8·6023	4·1983	0·01351	27·2029	0·11625	0·03676
75	5625	421875	8·6603	4·2172	0·01333	27·3861	0·11547	0·03651
76	5776	438976	8·7178	4·2358	0·01316	27·5681	0·11471	0·03627
77	5929	456533	8·7750	4·2543	0·01299	27·7489	0·11396	0·03604
78	6084	474552	8·8318	4·2727	0·01282	27·9285	0·11323	0·03581
79	6241	493039	8·8882	4·2908	0·01266	28·1069	0·11251	0·03558
80	6400	512000	8·9443	4·3089	0·01250	28·2843	0·11180	0·03536
81	6561	531441	9·0000	4·3267	0·01235	28·4605	0·11111	0·03514
82	6724	551368	9·0554	4·3445	0·01220	28·6356	0·11043	0·03492
83	6889	571787	9·1104	4·3621	0·01205	28·8097	0·10976	0·03471
84	7056	592704	9·1652	4·3795	0·01190	28·9828	0·10911	0·03450
85	7225	614125	9·2195	4·3968	0·01176	29·1548	0·10847	0·03430
86	7396	636056	9·2736	4·4140	0·01163	29·3258	0·10783	0·03410
87	7569	658503	9·3274	4·4310	0·01149	29·4958	0·10721	0·03390
88	7744	681472	9·3808	4·4480	0·01136	29·6648	0·10660	0·03371
89	7921	704969	9·4340	4·4647	0·01124	29·8329	0·10600	0·03352
90	8100	729000	9·4868	4·4814	0·01111	30·0000	0·10541	0·03333
91	8281	753571	9·5394	4·4979	0·01099	30·1662	0·10483	0·03315
92	8464	778688	9·5017	4·5144	0·01087	30·3315	0·10426	0·03297
93	8649	804357	9·6437	4·5307	0·01075	30·4959	0·10370	0·03279
94	8836	830584	9·6954	4·5468	0·01064	30·6594	0·10314	0·03262
95	9025	857375	9·7468	4·5629	0·01053	30·8221	0·10260	0·03244
96	9216	884736	9·7980	4·5789	0·01042	30·9839	0·10206	0·03227
97	9409	912673	9·8489	4·5947	0·01031	31·1448	0·10153	0·03211
98	9604	941192	9·8995	4·6104	0·01020	31·3050	0·10102	0·03194
99	9801	970299	9·9499	4·6261	0·01010	31·4643	0·10050	0·03178
100	10000	1000000	10·0000	4·6416	0·01000	31·6228	0·10000	0·03162

SQUARE ROOTS FROM 1 to 10

	0	1	2	3	4	5	6	7	8	9	Mean Differences		
											123	456	789
1·0	1·000	1·005	1·010	1·015	1·020	1·025	1·030	1·034	1·039	1·044	0 1 1	2 2 3	3 4 4
1·1	1·049	1·054	1·058	1·063	1·068	1·072	1·077	1·082	1·086	1·091	0 1 1	2 2 3	3 4 4
1·2	1·095	1·100	1·105	1·109	1·114	1·118	1·122	1·127	1·131	1·136	0 1 1	2 2 3	3 4 4
1·3	1·140	1·145	1·149	1·153	1·158	1·162	1·166	1·170	1·175	1·179	0 1 1	2 2 3	3 3 4
1·4	1·183	1·187	1·192	1·196	1·200	1·204	1·208	1·212	1·217	1·221	0 1 1	2 2 2	3 3 4
1·5	1·225	1·229	1·233	1·237	1·241	1·245	1·249	1·253	1·257	1·261	0 1 1	2 2 2	3 3 4
1·6	1·265	1·269	1·273	1·277	1·281	1·285	1·288	1·292	1·296	1·300	0 1 1	2 2 2	3 3 3
1·7	1·304	1·308	1·311	1·315	1·319	1·323	1·327	1·330	1·334	1·338	0 1 1	2 2 2	3 3 3
1·8	1·342	1·345	1·349	1·353	1·356	1·360	1·364	1·367	1·371	1·375	0 1 1	1 2 2	3 3 3
1·9	1·378	1·382	1·386	1·389	1·393	1·396	1·400	1·404	1·407	1·411	0 1 1	1 2 2	3 3 3
2·0	1·414	1·418	1·421	1·425	1·428	1·432	1·435	1·439	1·442	1·446	0 1 1	1 2 2	2 3 3
2·1	1·449	1·453	1·456	1·459	1·463	1·466	1·470	1·473	1·476	1·480	0 1 1	1 2 2	2 3 3
2·2	1·483	1·487	1·490	1·493	1·497	1·500	1·503	1·507	1·510	1·513	0 1 1	1 2 2	2 3 3
2·3	1·517	1·520	1·523	1·526	1·530	1·533	1·536	1·539	1·543	1·546	0 1 1	1 2 2	2 3 3
2·4	1·549	1·552	1·556	1·559	1·562	1·565	1·568	1·572	1·575	1·578	0 1 1	1 2 2	2 3 3
2·5	1·581	1·584	1·587	1·591	1·594	1·597	1·600	1·603	1·606	1·609	0 1 1	1 2 2	2 3 3
2·6	1·612	1·616	1·619	1·622	1·625	1·628	1·631	1·634	1·637	1·640	0 1 1	1 2 2	2 2 3
2·7	1·643	1·646	1·649	1·652	1·655	1·658	1·661	1·664	1·667	1·670	0 1 1	1 2 2	2 2 3
2·8	1·673	1·676	1·679	1·682	1·685	1·688	1·691	1·694	1·697	1·700	0 1 1	1 1 2	2 2 3
2·9	1·703	1·706	1·709	1·712	1·715	1·718	1·720	1·723	1·726	1·729	0 1 1	1 1 2	2 2 3
3·0	1·732	1·735	1·738	1·741	1·744	1·746	1·749	1·752	1·755	1·758	0 1 1	1 1 2	2 2 3
3·1	1·761	1·764	1·766	1·769	1·772	1·775	1·778	1·780	1·783	1·786	0 1 1	1 1 2	2 2 3
3·2	1·789	1·792	1·794	1·797	1·800	1·803	1·806	1·808	1·811	1·814	0 1 1	1 1 2	2 2 2
3·3	1·817	1·819	1·822	1·825	1·828	1·830	1·833	1·836	1·838	1·841	0 1 1	1 1 2	2 2 2
3·4	1·844	1·847	1·849	1·852	1·855	1·857	1·860	1·863	1·865	1·868	0 1 1	1 1 2	2 2 2
3·5	1·871	1·873	1·876	1·879	1·881	1·884	1·887	1·889	1·892	1·895	0 1 1	1 1 2	2 2 2
3·6	1·897	1·900	1·903	1·905	1·908	1·910	1·913	1·916	1·918	1·921	0 1 1	1 1 2	2 2 2
3·7	1·924	1·926	1·929	1·931	1·934	1·936	1·939	1·942	1·944	1·947	0 1 1	1 1 2	2 2 2
3·8	1·949	1·952	1·954	1·957	1·960	1·962	1·965	1·967	1·970	1·972	0 1 1	1 1 2	2 2 2
3·9	1·975	1·977	1·980	1·982	1·985	1·987	1·990	1·992	1·995	1·997	0 1 1	1 1 2	2 2 2
4·0	2·000	2·002	2·005	2·007	2·010	2·012	2·015	2·017	2·020	2·022	0 0 1	1 1 1	2 2 2
4·1	2·025	2·027	2·030	2·032	2·035	2·037	2·040	2·042	2·045	2·047	0 0 1	1 1 1	2 2 2
4·2	2·049	2·052	2·054	2·057	2·059	2·062	2·064	2·066	2·069	2·071	0 0 1	1 1 1	2 2 2
4·3	2·074	2·076	2·078	2·081	2·083	2·086	2·088	2·090	2·093	2·095	0 0 1	1 1 1	2 2 2
4·4	2·098	2·100	2·102	2·105	2·107	2·110	2·112	2·114	2·117	2·119	0 0 1	1 1 1	2 2 2
4·5	2·121	2·124	2·126	2·128	2·131	2·133	2·135	2·138	2·140	2·142	0 0 1	1 1 1	2 2 2
4·6	2·145	2·147	2·149	2·152	2·154	2·156	2·159	2·161	2·163	2·166	0 0 1	1 1 1	2 2 2
4·7	2·168	2·170	2·173	2·175	2·177	2·179	2·182	2·184	2·186	2·189	0 0 1	1 1 1	2 2 2
4·8	2·191	2·193	2·195	2·198	2·200	2·202	2·205	2·207	2·209	2·211	0 0 1	1 1 1	2 2 2
4·9	2·214	2·216	2·218	2·220	2·223	2·225	2·227	2·229	2·232	2·234	0 0 1	1 1 1	2 2 2
5·0	2·236	2·238	2·241	2·243	2·245	2·247	2·249	2·252	2·254	2·256	0 0 1	1 1 1	2 2 2
5·1	2·258	2·261	2·263	2·265	2·267	2·269	2·272	2·274	2·276	2·278	0 0 1	1 1 1	2 2 2
5·2	2·280	2·283	2·285	2·287	2·289	2·291	2·293	2·296	2·298	2·300	0 0 1	1 1 1	2 2 2
5·3	2·302	2·304	2·307	2·309	2·311	2·313	2·315	2·317	2·319	2·322	0 0 1	1 1 1	2 2 2
5·4	2·324	2·326	2·328	2·330	2·332	2·335	2·337	2·339	2·341	2·343	0 0 1	1 1 1	1 2 2

Square Roots from 1 to 10

	0	1	2	3	4	5	6	7	8	9	Mean Differences.								
											1	2	3	4	5	6	7	8	9
5·5	2·345	2·347	2·349	2·352	2·354	2·356	2·358	2·360	2·362	2·364	0	0	1	1	1	1	1	2	2
5·6	2·366	2·369	2·371	2·373	2·375	2·377	2·379	2·381	2·383	2·385	0	0	1	1	1	1	1	2	2
5·7	2·387	2·390	2·392	2·394	2·396	2·398	2·400	2·402	2·404	2·406	0	0	1	1	1	1	1	2	2
5·8	2·408	2·410	2·412	2·415	2·417	2·419	2·421	2·423	2·425	2·427	0	0	1	1	1	1	1	2	2
5·9	2·429	2·431	2·433	2·435	2·437	2·439	2·441	2·443	2·445	2·447	0	0	1	1	1	1	1	2	2
6·0	2·449	2·452	2·454	2·456	2·458	2·460	2·462	2·464	2·466	2·468	0	0	1	1	1	1	1	2	2
6·1	2·470	2·472	2·474	2·476	2·478	2·480	2·482	2·484	2·486	2·488	0	0	1		1	1	1	2	2
6·2	2·4900	2·492	2·494	2·496	2·498	2·500	2·502	2·504	2·506	2·508	0	0	1	1	1	1	1	2	2
6·3	2·510	2·512	2·514	2·516	2·518	2·520	2·522	2·524	2·526	2·528	0	0	1	1	1	1	1	2	2
6·4	2·530	2·532	2·534	2·536	2·538	2·540	2·542	2·544	2·546	2·548	0	0	1	1	1	1	1	2	2
6·5	2·550	2·551	2·553	2·555	2·557	2·559	2·561	2·563	2·565	2·567	0	0	1	1	1	1	1	2	2
6·6	2·569	2·571	2·573	2·575	2·577	2·579	2·581	2·583	2·585	2·587	0	0	1	1	1	1	1	2	2
6·7	2·588	2·590	2·592	2·594	2·596	2·598	2·600	2·602	2·604	2·606	0	0	1	1	1	1	1	2	2
6·8	2·608	2·610	2·612	2·613	2·615	2·617	2·619	2·621	2·623	2·625	0	0	1	1	1	1	1	2	2
6·9	2·627	2·629	2·631	2·632	2·634	2·636	2·638	2·640	2·642	2·644	0	0	1	1	1	1	1	2	2
7·0	2·646	2·648	2·650	2·651	2·653	2·655	2·657	2·659	2·661	2·663	0	0	1	1	1	1	1	2	2
7·1	2·665	2·666	2·668	2·670	2·672	2·674	2·676	2·678	2·680	2·681	0	0	1	1	1	1	1	1	2
7·2	2·683	2·685	2·687	2·689	2·691	2·693	2·694	2·696	2·698	2·700	0	0	1	1	1	1	1	1	2
7·3	2·702	2·704	2·706	2·707	2·709	2·711	2·713	2·715	2·717	2·718	0	0	1	1	1	1	1	1	2
7·4	2·720	2·722	2·724	2·726	2·728	2·729	2·731	2·733	2·735	2·737	0	0	1	1	1	1	1	1	2
7·5	2·739	2·740	2·742	2·744	2·746	2·748	2·750	2·751	2·753	2·755	0	0	1	1	1	1	1	1	2
7·6	2·757	2·759	2·760	2·762	2·764	2·766	2·768	2·769	2·771	2·773	0	0	1	1	1	1	1	1	2
7·7	2·775	2·777	2·778	2·780	2·782	2·784	2·786	2·787	2·789	2·791	0	0	1	1	1	1	1	1	2
7·8	2·793	2·795	2·796	2·798	2·800	2·802	2·804	2·805	2·807	2·809	0	0	1	1	1	1	1	1	2
7·9	2·811	2·812	2·814	2·816	2·818	2·820	2·821	2·823	2·825	2·827	0	0	1	1	1	1	1	1	2
8·0	2·828	2·830	2·832	2·834	2·835	2·837	2·839	2·841	2·843	2·844	0	0	1	1	1	1	1	1	2
8·1	2·846	2·848	2·850	2·851	2·853	2·855	2·857	2·858	2·860	2·862	0	0	1	1	1	1	1	1	2
8·2	2·864	2·865	2·867	2·869	2·871	2·872	2·874	2·876	2·877	2·879	0	0	1	1	1	1	1	1	2
8·3	2·881	2·883	2·884	2·886	2·888	2·890	2·891	2·893	2·895	2·897	0	0	1	1	1	1	1	1	2
8·4	2·898	2·900	2·902	2·903	2·905	2·907	2·909	2·910	2·912	2·914	0	0	1	1	1	1	1	1	2
8·5	2·915	2·917	2·919	2·921	2·922	2·924	2·926	2·927	2·929	2·931	0	0	1	1	1	1	1	1	2
8·6	2·933	2·934	2·936	2·938	2·939	2·941	2·943	2·944	2·946	2·948	0	0	1	1	1	1	1	1	2
8·7	2·950	2·951	2·953	2·955	2·956	2·958	2·960	2·961	2·963	2·965	0	0	1	1	1	1	1	1	2
8·8	2·966	2·968	2·970	2·972	2·973	2·975	2·977	2·978	2·980	2·982	0	0	1	1	1	1	1	1	2
8·9	2·983	2·985	2·987	2·988	2·990	2·992	2·993	2·995	2·997	2·998	0	0	1	1	1	1	1	1	2
9·0	3·000	3·002	3·003	3·005	3·007	3·008	3·010	3·012	3·013	3·015	0	0	0	1	1	1	1	1	1
9·1	3·017	3·018	3·020	3·022	3·023	3·025	3·027	3·028	3·030	3·032	0	0	0	1	1	1	1	1	1
9·2	3·033	3·035	3·036	3·038	3·040	3·041	3·043	3·045	3·046	3·048	0	0	0	1	1	1	1	1	1
9·3	3·050	3·051	3·053	3·055	3·056	3·058	3·059	3·061	3·063	3·064	0	0	0	1	1	1	1	1	1
9·4	3·066	3·068	3·069	3·071	3·072	3·074	3·076	3·077	3·079	3·081	0	0	0	1	1	1	1	1	1
9·5	3·082	3·084	3·085	3·087	3·089	3·090	3·092	3·094	3·095	3·097	0	0	0	1	1	1	1	1	1
9·6	3·098	3·100	3·102	3·103	3·105	3·106	3·108	3·110	3·111	3·113	0	0	0	1	1	1	1	1	1
9·7	3·114	3·116	3·118	3·119	3·121	3·122	3·124	3·126	3·127	3·129	0	0	0	1	1	1	1	1	1
9·8	3·130	3·132	3·134	3·135	3·137	3·138	3·140	3·142	3·143	3·145	0	0	0	1	1	1	1	1	1
9·9	3·146	3·148	3·150	3·151	3·153	3·154	3·156	3·158	3·159	3·161	0	0	0	1	1	1	1	1	1

SQUARE ROOTS FROM 10 to 100

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
10	3·162	3·178	3·194	3·209	3·225	3·240	3·256	3·271	3·286	3·302	2	3	5	6	8	9	11	12	14
11	3·317	3·332	3·347	3·362	3·376	3·391	3·406	3·421	3·435	3·450	1	3	4	6	[illegible]	9	10	12	13
12	3·464	3·479	3·493	3·507	3·521	3·536	3·550	3·564	3·578	3·592	1	3	4	6	7	8	10	11	13
13	3·606	3·619	3·633	3·647	3·661	3·674	3·688	3·701	3·715	3·728	1	3	4	5	7	8	10	11	12
14	3·742	3·755	3·768	3·782	3·795	3·808	3·821	3·834	3·847	3·860	1	3	4	5	7	8	9	11	12
15	3·873	3·886	3·899	3·912	3·924	3·937	3·950	3·962	3·975	3·987	1	3	4	5	6	8	9	10	11
16	4·000	4·012	4·025	4·037	4·050	4·062	4·074	4·087	4·099	4·111	1	2	4	5	6	7	9	10	11
17	4·123	4·135	4·147	4·159	4·171	4·183	4·195	4·207	4·219	4·231	1	2	4	5	6	7	8	10	11
18	4·243	4·254	4·266	4·278	4·290	4·301	4·313	4·324	4·336	4·347	1	2	3	5	6	7	8	9	10
19	4·359	4·370	4·382	4·393	4·405	4·416	4·427	4·438	4·450	4·461	1	2	3	5	6	7	8	9	10
20	4·472	4·483	4·494	4·506	4·517	4·528	4·539	4·550	4·561	4·572	1	2	3	4	6	7	8	9	10
21	4·583	4·593	4·604	4·615	4·626	4·637	4·648	4·658	4·669	4·680	1	2	3	4	5	6	8	9	10
22	4·690	4·701	4·712	4·722	4·733	4·743	4·754	4·764	4·775	4·785	1	2	3	4	5	6	7	8	9
23	4·796	4·806	4·817	4·827	4·837	4·848	4·858	4·868	4·879	4·889	1	2	3	4	5	6	7	8	9
24	4·899	4·909	4·919	4·930	4·940	4·950	4·960	4·970	4·980	4·990	1	2	3	4	5	6	7	8	9
25	5·000	5·010	5·020	5·030	5·040	5·050	5·060	5·070	5·079	5·089	1	2	3	4	5	6	7	8	9
26	5·099	5·109	5·119	5·128	5·138	5·148	5·158	5·167	5·177	5·187	1	2	3	4	5	6	7	8	9
27	5·196	5·206	5·215	5·225	5·235	5·244	5·254	5·263	5·273	5·282	1	2	3	4	5	6	7	8	9
28	5·292	5·301	5·310	5·320	5·329	5·339	5·348	5·357	5·367	5·376	1	2	3	4	5	6	7	7	8
29	5·385	5·394	5·404	5·413	5·422	5·431	5·441	5·450	5·459	5·468	1	2	3	4	5	5	6	7	8
30	5·477	5·486	5·495	5·505	5·514	5·523	5·532	5·541	5·550	5·559	1	2	3	4	4	5	6	7	8
31	5·568	5·577	5·586	5·595	5·604	5·612	5·621	5·630	5·639	5·648	1	2	3	3	4	5	6	7	8
32	5·657	5·666	5·675	5·683	5·692	5·701	5·710	5·718	5·727	5·736	1	2	3	3	4	5	6	7	8
33	5·745	5·753	5·762	5·771	5·779	5·788	5·797	5·805	5·814	5·822	1	2	3	3	4	5	6	7	8
34	5·831	5·840	5·848	5·857	5·865	5·874	5·882	5·891	5·899	5·908	1	2	3	3	4	5	6	7	8
35	5·916	5·925	5·933	5·941	5·950	5·958	5·967	5·975	5·983	5·992	1	2	2	3	4	5	6	7	8
36	6·000	6·008	6·017	6·025	6·033	6·042	6·050	6·058	6·066	6·075	1	2	2	3	4	5	6	7	7
37	6·083	6·091	6·099	6·107	6·116	6·124	6·132	6·140	6·148	6·156	1	2	2	3	4	5	6	7	7
38	6·164	6·173	6·181	6·189	6·197	6·205	6·213	6·221	6·229	6·237	1	2	2	3	4	5	6	6	7
39	6·245	6·253	6·261	6·269	6·277	6·285	6·293	6·301	6·309	6·317	1	2	2	3	4	5	6	6	7
40	6·325	6·332	6·340	6·348	6·356	6·364	6·372	6·380	6·387	6·395	1	2	2	3	4	5	6	6	7
41	6·403	6·411	6·419	6·427	6·434	6·442	6·450	6·458	6·465	6·473	1	2	2	3	4	5	5	6	7
42	6·481	6·488	6·496	6·504	6·512	6·519	6·527	6·535	6·542	6·550	1	2	2	3	4	5	5	6	7
43	6·557	6·565	6·573	6·580	6·588	6·595	6·603	6·611	6·618	6·626	1	2	2	3	4	5	5	6	7
44	6·633	6·641	6·648	6·656	6·663	6·671	6·678	6·686	6·693	6·701	1	2	2	3	4	5	5	6	7
45	6·708	6·716	6·723	6·731	6·738	6·745	6·753	6·760	6·768	6·775	1	1	2	3	4	4	5	6	7
46	6·782	6·790	6·797	6·804	6·812	6·819	6·826	6·834	6·841	6·848	1	1	2	3	4	4	5	6	7
47	6·856	6·863	6·870	6·877	6·885	6·892	6·899	6·907	6·914	6·921	1	1	2	3	4	4	5	6	7
48	6·928	6·935	6·943	6·950	6·957	6·964	6·971	6·979	6·986	6·993	1	1	2	3	4	4	5	6	6
49	7·000	7·007	7·014	7·021	7·029	7·036	7·043	7·050	7·057	7·064	1	1	2	3	4	4	5	6	6
50	7·071	7·078	7·085	7·092	7·099	7·106	7·113	7·120	7·127	7·134	1	1	2	3	4	4	5	6	6
51	7·141	7·148	7·155	7·162	7·169	7·176	7·183	7·190	7·197	7·204	1	1	2	3	4	4	5	6	6
52	7·211	7·218	7·225	7·232	7·239	7·246	7·253	7·259	7·266	7·273	1	1	2	3	3	4	5	6	6
53	7·280	7·287	7·294	7·301	7·308	7·314	7·321	7·328	7·335	7·342	1	1	2	3	3	4	5	5	6
54	7·348	7·355	7·362	7·369	7·376	7·382	7·389	7·396	7·403	7·409	1	1	2	3	3	4	5	5	6

SQUARE ROOTS. FROM 10 TO 100

	0	1	2	3	4	5	6	7	8	9	Mean Differences		
											123	456	789
55	7·416	7·423	7·430	7·436	7·443	7·450	7·457	7·463	7·470	7·477	1 1 2	3 3 4	5 5 6
56	7·483	7·490	7·497	7·503	7·510	7·517	7·523	7·530	7·537	7·543	1 1 2	3 3 4	5 5 6
57	7·550	7·556	7·563	7·570	7·576	7·583	7·589	7·596	7·603	7·609	1 1 2	3 3 4	5 5 6
58	7·616	7·622	7·629	7·635	7·642	7·649	7·655	7·662	7·668	7·675	1 1 2	3 3 4	5 5 6
59	7·681	7·688	7·694	7·701	7·707	7·714	7·720	7·727	7·733	7·740	1 1 2	3 3 4	4 5 6
60	7·746	7·752	7·759	7·765	7·772	7·778	7·785	7·791	7·797	7·804	1 1 2	3 3 4	4 5 6
61	7·810	7·817	7·823	7·829	7·836	7·842	7·849	7·855	7·861	7·868	1 1 2	3 3 4	4 5 6
62	7·874	7·880	7·887	7·893	7·899	7·906	7·912	7·918	7·925	7·931	1 1 2	3 3 4	4 5 6
63	7·937	7·944	7·950	7·956	7·962	7·969	7·975	7·981	7·987	7·994	1 1 2	3 3 4	4 5 6
64	8·000	8·006	8·012	8·019	8·025	8·031	8·037	8·044	8·050	8·056	1 1 2	2 3 4	4 5 6
65	8·062	8·068	8·075	8·081	8·087	8·093	8·099	8·106	8·112	8·118	1 1 2	2 3 4	4 5 6
66	8·124	8·130	8·136	8·142	8·149	8·155	8·161	8·167	8·173	8·179	1 1 2	2 3 4	4 5 5
67	8·185	8·191	8·198	8·204	8·210	8·216	8·222	8·228	8·234	8·240	1 1 2	2 3 4	4 5 5
68	8·246	8·252	8·258	8·264	8·270	8·276	8·283	8·289	8·295	8·301	1 1 2	2 3 4	4 5 5
69	8·307	8·313	8·319	8·325	8·331	8·337	8·343	8·349	8·355	8·361	1 1 2	2 3 4	4 5 5
70	8·367	8·373	8·379	8·385	8·390	8·396	8·402	8·408	8·414	8·420	1 1 2	2 3 4	4 5 5
71	8·426	8·432	8·438	8·444	8·450	8·456	8·462	8·468	8·473	8·479	1 1 2	2 3 4	4 5 5
72	8·485	8·491	8·497	8·503	8·509	8·515	8·521	8·526	8·532	8·538	1 1 2	2 3 3	4 5 5
73	8·544	8·550	8·556	8·562	8·567	8·573	8·579	8·585	8·591	8·597	1 1 2	2 3 3	4 5 5
74	8·602	8·608	8·614	8·620	8·626	8·631	8·637	8·643	8·649	8·654	1 1 2	2 3 3	4 5 5
75	8·660	8·666	8·672	8·678	8·683	8·689	8·695	8·701	8·706	8·712	1 1 2	2 3 3	4 5 5
76	8·718	8·724	8·729	8·735	8·741	8·746	8·752	8·758	8·764	8·769	1 1 2	2 3 3	4 5 5
77	8·775	8·781	8·786	8·792	8·798	8·803	8·809	8·815	8·820	8·826	1 1 2	2 3 3	4 4 5
78	8·832	8·837	8·843	8·849	8·854	8·860	8·866	8·871	8·877	8·883	1 1 2	2 3 3	4 4 5
79	8·888	8·894	8·899	8·905	8·911	8·916	8·922	8·927	8·933	8·939	1 1 2	2 3 3	4 4 5
80	8·944	8·950	8·955	8·961	8·967	8·972	8·978	8·983	8·989	8·994	1 1 2	2 3 3	4 4 5
81	9·000	9·006	9·011	9·017	9·022	9·028	9·033	9·039	9·044	9·050	1 1 2	2 3 3	4 4 5
82	9·055	9·061	9·066	9·072	9·077	9·083	9·088	9·094	9·099	9·105	1 1 2	2 3 3	4 4 5
83	9·110	9·116	9·121	9·127	9·132	9·138	9·143	9·149	9·154	9·160	1 1 2	2 3 3	4 4 5
84	9·165	9·171	9·176	9·182	9·187	9·192	9·198	9·203	9·209	9·214	1 1 2	2 3 3	4 4 5
85	9·220	9·225	9·230	9·236	9·241	9·247	9·252	9·257	9·263	9·268	1 1 2	2 3 3	4 4 5
86	9·274	9·279	9·284	9·290	9·295	9·301	9·306	9·311	9·317	9·322	1 1 2	2 3 3	4 4 5
87	9·327	9·333	9·338	9·343	9·349	9·354	9·359	9·365	9·370	9·375	1 1 2	2 3 3	4 4 5
88	9·381	9·386	9·391	9·397	9·402	9·407	9·413	9·418	9·423	9·429	1 1 2	2 3 3	4 4 5
89	9·434	9·439	9·445	9·450	9·455	9·460	9·466	9·471	9·476	9·482	1 1 2	2 3 3	4 4 5
90	9·487	9·492	9·497	9·503	9·508	9·513	9·518	9·524	9·529	9·534	1 1 2	2 3 3	4 4 5
91	9·539	9·545	9·550	9·555	9·560	9·566	9·571	9·576	9·581	9·586	1 1 2	2 3 3	4 4 5
92	9·592	9·597	9·602	9·607	9·612	9·618	9·623	9·628	9·633	9·638	1 1 2	2 3 3	4 4 5
93	9·644	9·649	9·654	9·659	9·664	9·670	9·675	9·680	9·685	9·690	1 1 2	2 3 3	4 4 5
94	9·695	9·701	9·706	9·711	9·716	9·721	9·726	9·731	9·737	9·742	1 1 2	2 3 3	4 4 5
95	9·747	9·752	9·757	9·762	9·767	9·772	9·778	9·783	9·788	9·793	1 1 2	2 3 3	4 4 5
96	9·798	9·803	9·808	9·813	9·818	9·823	9·829	9·834	9·839	9·844	1 1 2	2 3 3	4 4 5
97	9·849	9·854	9·859	9·864	9·869	9·874	9·879	9·884	9·889	9·894	1 1 1	2 3 3	4 4 5
98	9·899	9·905	9·910	9·915	9·920	9·925	9·930	9·935	9·940	9·945	0 1 1	2 2 3	3 4 4
99	9·950	9·955	9·960	9·965	9·970	9·975	9·980	9·985	9·990	9·995	0 1 1	2 2 3	3 4 4

RECIPROCALS OF NUMBERS FROM 1 to 10

[Numbers in difference columns to be subtracted, not added.]

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
1·0	1·000	9901	9804	9709	9615	9524	9434	9346	9259	9174									
1·1	·9091	9009	8929	8850	8772	8696	8621	8547	8475	8403									
1·2	·8333	8264	8197	8130	8065	8000	7937	7874	7813	7752									
1·3	·7692	7634	7576	7519	7463	7407	7353	7299	7246	7194									
1·4	·7143	7092	7042	6993	6944	6897	6849	6803	6757	6711	5	10	14	19	24	29	33	38	43
1·5	·6667	6623	6579	6536	6494	6452	6410	6369	6329	6289	4	8	13	17	21	25	29	33	38
1·6	·0250	6211	6173	6135	6098	6061	6024	5988	5952	5917	4	7	11	15	18	22	26	29	33
1·7	·5882	5848	5814	5780	5747	5714	5682	5650	5618	5587	3	6	10	13	16	20	23	26	29
1·8	·5556	5525	5495	5464	5435	5405	5376	5348	5319	5291	3	6	9	12	15	17	20	23	26
1·9	·5263	5236	5208	5181	5155	5128	5102	5076	5051	5025	3	5	8	11	13	16	18	21	24
2·0	·5000	4975	4950	4926	4902	4878	4854	4831	4808	4785	2	5	7	10	12	14	17	19	21
2·1	·4762	4739	4717	4695	4673	4651	4630	4608	4587	4566	2	4	7	9	11	13	15	17	20
2·2	·4545	4525	4505	4484	4464	4444	4425	4405	4386	4367	2	4	6	8	10	12	14	16	18
2·3	·4348	4329	4310	4292	4274	4255	4237	4219	4202	4184	2	4	5	7	9	11	13	14	16
2·4	·4167	4149	4132	4115	4098	4082	4065	4049	4032	4016	2	3	5	7	8	10	12	13	15
2·5	·4000	3984	3968	3953	3937	3922	3906	3891	3876	3861	2	3	5	6	8	9	11	12	14
2·6	·3846	3831	3817	3802	3788	3774	3759	3745	3731	3717	1	3	4	6	7	8	10	11	13
2·7	·3704	3690	3676	3663	3650	3636	3623	3610	3597	3584	1	3	4	5	7	8	9	11	12
2·8	·3571	3559	3546	3534	3521	3509	3497	3484	3472	3460	1	2	4	5	6	7	9	10	11
2·9	·3448	3436	3425	3413	3401	3390	3378	3367	3356	3344	1	2	3	5	6	7	8	9	10
3·0	·3333	3322	3311	3300	3289	3279	3268	3257	3247	3236	1	2	3	4	5	6	7	9	10
3·1	·3226	3215	3205	3195	3185	3175	3165	3155	3145	3135	1	2	3	4	5	6	7	8	9
3·2	·3125	3115	3106	3096	3086	3077	3067	3058	3049	3040	1	2	3	4	5	6	7	8	9
3·3	·3030	3021	3012	3003	2994	2985	2976	2967	2959	2950	1	2	3	4	4	5	6	7	8
3·4	·2941	2933	2924	2915	2907	2899	2890	2882	2874	2865	1	2	3	3	4	5	6	7	8
3·5	·2857	2849	2841	2833	2825	2817	2809	2801	2793	2786	1	2	2	3	4	5	6	6	7
3·6	·2778	2770	2762	2755	2747	2740	2732	2725	2717	2710	1	2	2	3	4	5	5	6	7
3·7	·2703	2695	2688	2681	2674	2667	2660	2653	2646	2639	1	1	2	3	4	4	5	6	6
3·8	·2632	2625	2618	2611	2604	2597	2591	2584	2577	2571	1	1	2	3	3	4	5	5	6
3·9	·2564	2558	2551	2545	2538	2532	2525	2519	2513	2506	1	1	2	3	3	4	4	5	6
4·0	·2500	2494	2488	2481	2475	2469	2463	2457	2451	2445	1	1	2	2	3	4	4	5	5
4·1	·2439	2433	2427	2421	2415	2410	2404	2398	2392	2387	1	1	2	2	3	3	4	5	5
4·2	·2381	2375	2370	2364	2358	2353	2347	2342	2336	2331	1	1	2	2	3	3	4	4	5
4·3	·2326	2320	2315	2309	2304	2299	2294	2288	2283	2278	1	1	2	2	3	3	4	4	5
4·4	·2273	2268	2262	2257	2252	2247	2242	2237	2232	2227	1	1	2	2	3	3	4	4	5
4·5	·2222	2217	2212	2208	2203	2198	2193	2188	2183	2179	0	1	1	2	2	3	3	4	4
4·6	·2174	2169	2165	2160	2155	2151	2146	2141	2137	2132	0	1	1	2	2	3	3	4	4
4·7	·2128	2123	2119	2114	2110	2105	2101	2096	2092	2088	0	1	1	2	2	3	3	4	4
4·8	·2083	2079	2075	2070	2066	2062	2058	2053	2049	2045	0	1	1	2	2	3	3	3	4
4·9	·2041	2037	2033	2028	2024	2020	2016	2012	2008	2004	0	1	1	2	2	2	3	3	4
5·0	·2000	1996	1992	1988	1984	1980	1976	1972	1969	1965	0	1	1	2	2	2	3	3	4
5·1	·1961	1957	1953	1949	1946	1942	1938	1934	1931	1927	0	1	1	2	2	2	3	3	3
5·2	·1923	1919	1916	1912	1908	1905	1901	1898	1894	1890	0	1	1	1	2	2	3	3	3
5·3	·1887	1883	1880	1876	1873	1869	1866	1862	1859	1855	0	1	1	1	2	2	2	3	3
5·4	·1852	1848	1845	1842	1838	1835	1832	1828	1825	1821	0	1	1	1	2	2	2	3	3

RECIPROCALS OF NUMBERS FROM 1 to 10

[Numbers in difference columns to be subtracted, not added.]

	0	1	2	3	4	5	6	7	8	9	Mean Differences		
											123	456	789
5·5	·1818	1815	1812	1808	1805	1802	1799	1795	1792	1789	0 1 1	1 2 2	2 3 3
5·6	·1786	1783	1779	1776	1773	1770	1767	1764	1761	1757	0 1 1	1 2 2	2 3 3
5·7	·1754	1751	1748	1745	1742	1739	1736	1733	1730	1727	0 1 1	1 1 2	2 2 3
5·8	·1724	1721	1718	1715	1712	1709	1706	1704	1701	1698	0 1 1	1 1 2	2 2 3
5·9	·1695	1692	1689	1686	1684	1681	1678	1675	1672	1669	0 1 1	1 1 2	2 2 3
6·0	·1667	1664	1661	1658	1656	1653	1650	1647	1645	1642	0 1 1	1 1 2	2 2 3
6·1	·1639	1637	1634	1631	1629	1626	1623	1621	1618	1616	0 1 1	1 1 2	2 2 2
6·2	·1613	1610	1608	1605	1603	1600	1597	1595	1592	1590	0 1 1	1 1 2	2 2 2
6·3	·1587	1585	1582	1580	1577	1575	1572	1570	1567	1565	0 0 1	1 1 1	2 2 2
6·4	·1562	1560	1558	1555	1553	1550	1548	1546	1543	1541	0 0 1	1 1 1	2 2 2
6·5	·1538	1536	1534	1531	1529	1527	1524	1522	1520	1517	0 0 1	1 1 1	2 2 2
6·6	·1515	1513	1511	1508	1506	1504	1502	1499	1497	1495	0 0 1	1 1 1	2 2 2
6·7	·1493	1490	1488	1486	1484	1481	1479	1477	1475	1473	0 0 1	1 1 1	2 2 2
6·8	·1471	1468	1466	1464	1462	1460	1458	1456	1453	1451	0 0 1	1 1 1	2 2 2
6·9	·1449	1447	1445	1443	1441	1439	1437	1435	1433	1431	0 0 1	1 1 1	2 2 2
7·0	·1429	1427	1425	1422	1420	1418	1416	1414	1412	1410	0 0 1	1 1 1	1 2 2
7·1	·1408	1406	1404	1403	1401	1399	1397	1395	1393	1391	0 0 1	1 1 1	1 2 2
7·2	·1389	1387	1385	1383	1381	1379	1377	1376	1374	1372	0 0 1	1 1 1	1 2 2
7·3	·1370	1368	1366	1364	1362	1361	1359	1357	1355	1353	0 0 1	1 1 1	1 2 2
7·4	·1351	1350	1348	1346	1344	1342	1340	1339	1337	1335	0 0 1	1 1 1	1 1 2
7·5	·1333	1332	1330	1328	1326	1325	1323	1321	1319	1318	0 0 1	1 1 1	1 1 2
7·6	·1316	1314	1312	1311	1309	1307	1305	1304	1302	1300	0 0 1	1 1 1	1 1 2
7·7	·1299	1297	1295	1294	1292	1290	1289	1287	1285	1284	0 0 0	1 1 1	1 1 1
7·8	·1282	1280	1279	1277	1276	1274	1272	1271	1269	1267	0 0 0	1 1 1	1 1 1
7·9	·1266	1264	1263	1261	1259	1258	1256	1255	1253	1252	0 0 0	1 1 1	1 1 1
8·0	·1250	1248	1247	1245	1244	1242	1241	1239	1238	1236	0 0 0	1 1 1	1 1 1
8·1	·1235	1233	1232	1230	1229	1227	1225	1224	1222	1221	0 0 0	1 1 1	1 1 1
8·2	·1220	1218	1217	1215	1214	1212	1211	1209	1208	1206	0 0 0	1 1 1	1 1 1
8·3	·1205	1203	1202	1200	1199	1198	1196	1195	1193	1192	0 0 0	1 1 1	1 1 1
8·4	·1190	1189	1188	1186	1185	1183	1182	1181	1179	1178	0 0 0	1 1 1	1 1 1
8·5	·1176	1175	1174	1172	1171	1170	1168	1167	1166	1164	0 0 0	1 1 1	1 1 1
8·6	·1163	1161	1160	1159	1157	1156	1155	1153	1152	1151	0 0 0	1 1 1	1 1 1
8·7	·1149	1148	1147	1145	1144	1143	1142	1140	1139	1138	0 0 0	1 1 1	1 1 1
8·8	·1136	1135	1134	1133	1131	1130	1129	1127	1126	1125	0 0 0	1 1 1	1 1 1
8·9	·1124	1122	1121	1120	1119	1117	1116	1115	1114	1112	0 0 0	1 1 1	1 1 1
9·0	·1111	1110	1109	1107	1106	1105	1104	1103	1101	1100	0 0 0	1 1 1	1 1 1
9·1	·1099	1098	1096	1095	1094	1093	1092	1090	1089	1088	0 0 0	0 1 1	1 1 1
9·2	·1087	1086	1085	1083	1082	1081	1080	1079	1078	1076	0 0 0	0 1 1	1 1 1
9·3	·1075	1074	1073	1072	1071	1070	1068	1067	1066	1065	0 0 0	0 1 1	1 1 1
9·4	·1064	1063	1062	1060	1059	1058	1057	1056	1055	1054	0 0 0	0 1 1	1 1 1
9·5	·1053	1052	1050	1049	1048	1047	1046	1045	1044	1043	0 0 0	0 1 1	1 1 1
9·6	·1042	1041	1039	1038	1037	1036	1035	1034	1033	1032	0 0 0	0 1 1	1 1 1
9·7	·1031	1030	1029	1028	1027	1026	1025	1024	1022	1021	0 0 0	0 1 1	1 1 1
9·8	·1020	1019	1018	1017	1016	1015	1014	1013	1012	1011	0 0 0	0 1 1	1 1 1
9·9	·1010	1009	1008	1007	1006	1005	1004	1003	1002	1001	0 0 0	0 0 1	1 1 1

TABLE Ordinates and Areas of the Standard Normal Curve

(In terms of σ units)

$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate
00	.0000	.3989	.50	.1915	.3521	1.00	.3413	.2420
.01	.0040	.3989	.51	.1950	.3503	1.01	.3438	.2396
.02	.0080	.3989	.52	.1985	.3485	1.02	.3461	.2371
.03	.0120	.3988	.53	.2019	.3467	1.03	.3485	.2347
.04	.0160	.3986	.54	.2054	.3448	1.04	.3508	.2323
.05	.0199	.3984	.55	.2088	.3429	1.05	.3531	.2299
.06	.0239	.3982	.56	.2123	.3410	1.06	.3554	.2275
.07	.0279	.3980	.57	.2157	.3291	1.07	.3577	.2251
.08	.0319	.3977	.58	.2190	.3372	1.08	.3599	.2227
.09	.0359	.3973	.59	.2224	.3352	1.09	.3621	.2203
.10	.0398	.3970	.60	.2257	.3332	1.10	.3643	.2179
.11	.0438	.3965	.61	.2291	.3312	1.11	.3665	.2155
.12	.0478	.3961	.62	.2324	.3292	1.12	.3686	.2131
.13	.0517	.3956	.63	.2357	.3271	1.13	.3708	.2107
.14	.0557	.3951	.64	.2389	.3251	1.14	.3729	.2083
.15	.0596	.3945	.65	.2422	.3230	1.15	.3749	.2059
.16	.0636	.3939	.66	.2454	.3209	1.16	.3770	.2036
.17	.0675	.3932	.67	.2486	.3187	1.17	.3790	.2012
.18	.0714	.3925	.68	.2517	.3166	1.18	.3810	.1989
.19	.0753	.3918	.69	.2549	.3144	1.19	.3830	.1965
.20	.0793	.3910	.70	.2580	.3123	1.20	.3849	.1942
.21	.0832	.3902	.71	.2611	.3101	1.21	.3869	.1919
.22	.0871	.3894	.72	.2642	.3079	1.22	.3888	.1895
.23	.0910	.3885	.73	.2673	.3056	1.23	.3907	.1872
.24	.0948	.3876	.74	.2703	.3034	1.24	.3925	.1849
.25	.0987	.3867	.75	.2734	.3011	1.25	.3944	.1826
.26	.1026	.3857	.76	.2764	.2989	1.26	.3962	.1804
.27	.1064	.3847	.77	.2794	.2966	1.27	.3980	.1781
.28	.1103	.3836	.78	.2823	.2943	1.28	.3997	.1758
.29	.1141	.3825	.79	.2852	.2920	1.29	.4015	.1736
.30	.1179	.3814	80	.2881	.2897	1.30	.4032	.1714
.31	.1217	.3802	.81	.2910	.2874	1.31	.4049	.1691
.32	.1255	.3790	.82	.2939	.2850	1.32	.4066	.1669
.33	.1293	.3778	.83	.2967	.2827	1.33	.4082	.1647
.34	.1331	.3765	.84	.2995	.2803	1.34	.4099	.1626
.35	.1368	.3752	.85	.3023	.2780	1.35	.4115	.1604
.36	.1406	.3739	.86	.3051	.2756	1.36	.4131	.1582
.37	.1443	.3725	.87	.3078	.2732	1.37	.4147	.1561
.38	.1480	.3712	.88	.3106	.2709	1.38	.4162	.1539
.39	.1517	.3697	.89	.3133	.2685	1.39	.4177	.1518
.40	.1554	.3683	.90	.3159	.2661	1.40	.4192	.1497
.41	.1591	.3668	.91	.3186	.2637	1.41	.4207	.1476
.42	.1628	.3653	.92	.3212	.2613	1.42	.4222	.1456
.43	.1664	.3637	.93	.3238	.2589	1.43	.4236	.1435
.44	.1700	.3621	.94	.3264	.2565	1.44	.4251	.1415
.45	.1736	.3605	.95	.3289	.2541	1.45	.4265	.1394
.46	.1772	.3589	.96	.3315	.2516	1.46	.4279	.1374
.47	.1808	.3572	.97	.3340	.2402	1.47	.4292	.1354
.48	.1844	.3555	.98	.3365	.2468	1.48	.4306	.1334
.49	.1879	.3538	.99	.3389	.2444	1.49	.4319	.1315
.50	.1915	.3521	1.00	.3413	.2420	1.50	.4332	.1295

TABLE I Ordinates and Areas of the Standard Normal Curve. *Concluded*

(In terms of σ units)

$\frac{z}{\sigma}$	Area	Ordinate	$\frac{z}{\sigma}$	Area	Ordinate	$\frac{z}{\sigma}$	Area	Ordinate
1.50	.4332	.1295	2.00	.4772	.0540	2.50	.4938	.0175
1.51	.4345	.1276	2.01	.4778	.0529	2.51	.4940	.0171
1.52	.4357	.1257	2.02	.4783	.0519	2.52	.4941	.0167
1.53	.4370	.1238	2.03	.4788	.0508	2.53	.4943	.0163
1.54	.4382	.1219	2.04	.4793	.0498	2.54	.4945	.0158
1.55	.4394	.1200	2.05	.4798	.0488	2.55	.4946	.0154
1.56	.4406	.1182	2.06	.4803	.0478	2.56	.4948	.0151
1.57	.4418	.1163	2.07	.4808	.0468	2.57	.4949	.0147
1.58	.4429	.1145	2.08	.4812	.0459	2.58	.4951	.0143
1.59	.4441	.1127	2.09	.4817	.0449	2.59	.4952	.0139
1.60	.4452	.1109	2.10	.4821	.0440	2.60	.4953	.0136
1.61	.4463	.1092	2.11	.4826	.0431	2.61	.4955	.0132
1.62	.4474	.1074	2.12	.4830	.0422	2.62	.4956	.0129
1.63	.4484	.1057	2.13	.4834	.0413	2.63	.4957	.0126
1.64	.4495	.1040	2.14	.4838	.0404	2.64	.4959	.0122
1.65	.4505	.1023	2.15	.4842	.0395	2.65	.4960	.0119
1.66	.4515	.1006	2.16	.4846	.0387	2.66	.4961	.0116
1.67	.4525	.0989	2.17	.4850	.0379	2.67	.4962	.0113
1.68	.4535	.0973	2.18	.4854	.0371	2.68	.4963	.0110
1.69	.4545	.0957	2.19	.4857	.0363	2.69	.4964	.0107
1.70	.4554	.0940	2.20	.4861	.0355	2.70	.4965	.0104
1.71	.4564	.0925	2.21	.4864	.0347	2.71	.4966	.0101
1.72	.4573	.0909	2.22	.4868	.0339	2.72	.4967	.0099
1.73	.4582	.0893	2.23	.4871	.0332	2.73	.4968	.0096
1.74	.4591	.0878	2.24	.4875	.0325	2.74	.4969	.0093
1.75	.4599	.0863	2.25	.4878	.0317	2.75	.4970	.0091
1.76	.4608	.0848	2.26	.4881	.0310	2.76	.4971	.0088
1.77	.4616	.0833	2.27	.4884	.0303	2.77	.4972	.0086
1.78	.4625	.0818	2.28	.4887	.0297	2.78	.4973	.0084
1.79	.4633	.0804	2.29	.4890	.0290	2.79	.4974	.0081
1.80	.4641	.0790	2.30	.4893	.0283	2.80	.4974	.0079
1.81	.4649	.0775	2.31	.4896	.0277	2.81	.4975	.0077
1.82	.4656	.0761	2.32	.4898	.0270	2.82	.4976	.0075
1.83	.4664	.0748	2.33	.4901	.0264	2.83	.4977	.0073
1.84	.4671	.0734	2.34	.4904	.0258	2.84	.4977	.0071
1.85	.4678	.0721	2.35	.4906	.0252	2.85	.4978	.0069
1.86	.4686	.0707	2.36	.4909	.0246	2.86	.4979	.0067
1.87	.4693	.0694	2.37	.4911	.0241	2.87	.4979	.0065
1.88	.4699	.0681	2.38	.4913	.0235	2.88	.4980	.0063
1.89	.4706	.0669	2.39	.4916	.0229	2.89	.4981	.0061
1.90	.4713	.0656	2.40	.4918	.0224	2.90	.4981	.0060
1.91	.4719	.0644	2.41	.4920	.0219	2.91	.4982	.0058
1.92	.4726	.0632	2.42	.4922	.0213	2.92	.4982	.0056
1.93	.4732	.0620	2.43	.4925	.0208	2.93	.4983	.0055
1.94	.4738	.0608	2.44	.4927	.0203	2.94	.4984	.0053
1.95	.4744	.0596	2.45	.4929	.0198	2.95	.4984	.0051
1.96	.4750	.0584	2.46	.4931	.0194	2.96	.4985	.0050
1.97	.4756	.0573	2.47	.4932	.0189	2.97	.4985	.0048
1.98	.4761	.0562	2.48	.4934	.0184	2.98	.4986	.0047
1.99	.4767	.0551	2.49	.4936	.0180	2.99	.4986	.0046
2.00	.4772	.0540	2.50	.4938	.0175	3.00	.4987	.0044

TABLE OF X^2 AND "*t*"

Degrees of freedom	X		"*t*"	
	PROBABILITY		PROBABILITY	
	.05	.01	.05	.01
1	3.84	6.64	12.71	63.66
2	5.99	9.21	4.30	9.92
3	7.82	11.34	3.18	5.84
4	9.49	13.28	2.78	4.60
5	11.07	15.09	2.57	4.03
6	12.59	16.81	2.45	3.71
7	14.07	18.48	2.36	3.50
8	15.51	20.09	2.31	3.36
9	16.92	21.67	2.26	3.25
10	18.31	23.21	2.23	3.17
11	19.68	24.72	2.20	3.11
12	21.03	26.22	2.18	3.06
13	22.36	27.69	2.16	3.01
14	23.68	29.14	2.14	2.98
15	25.00	30.58	2.13	2.95
16	26.30	32.00	2.12	2.92
17	27.59	33.41	2.11	2.90
18	28.87	34.80	2.10	2.88
19	30.14	36.19	2.09	2.86
20	31.41	37.57	2.09	2.84
21	32.67	38.93	2.08	2.83
22	33.92	40.29	2.07	2.82
23	35.17	41.67	2.07	2.81
24	36.42	42.98	2.06	2.80
25	37.65	44.31	2.06	2.79

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road